DOCUMENT RESUME

ED 479 869 TM 035 116

TITLE	CATS 2002 Interpretive Guide Detailed Information about How To Use Your Score Reports (Version 1.02).
INSTITUTION	Kentucky State Dept. of Education, Frankfort.
PUB DATE	2003-01-03
NOTE	98p.
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	EDRS Price MF01/PC04 Plus Postage.
DESCRIPTORS	Accountability; Elementary Secondary Education; *Reports; *Scores; State Programs; *Test Results; Test Use; *Testing Programs
IDENTIFIERS	*Kentucky

ABSTRACT

This document provides detailed information on how to interpret and use the results provided by the Kentucky Core Content Test (KCCT), Writing Portfolio, Norm-referenced Test, and other components of the Commonwealth Accountability Testing System (CATS) administered during the 2001-2002 school year. As required by statute, these reports are received by school districts on or before September 15 each year. Individual pages of the Kentucky Performance Report and three separate student level reports are explained in detail. This interpretive guide contains detailed information about the four performance levels that are the heart and soul of the CATS: Novice, Apprentice, Proficient, and Distinguished. How these performance levels were developed is explained. The accommodations and modifications made to individual tests in the assessment system are described, and each of the tests is also described. Scoring and scaling are also reviewed for each of the tests that make up the CATS. One appendix outlines the cut points for the performance levels, and the second appendix contains a glossary for the CATS. The third appendix is a series of questions commonly asked about the CATS, with answers. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 479 869

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

a. Arrastia

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CATS 2002 Interpretive Guide
Detailed Information About How to Use Your Score Reports

PROFICIENCY AND BEYOND

Kentucky Department
of Education

TM035116

KENTUCKY DEPARTMENT OF EDUCATION
Gene Wilhoit, Commissioner

(Version 1.02)

CATS 2002 Interpretive Guide
Detailed Information About How to Use Your Score Reports

Overview

This document gives detailed information on how to interpret and use the results provided by the Kentucky Core Content Test (KCCT), Writing Portfolio, Norm-Referenced Test and other components of the Commonwealth Accountability Testing System (CATS) administered during the 2001-2002 school year. As required in statute, these reports are received by school districts on or before September 15th each year. The following individual pages of the Kentucky Performance Report (KPR) and three separate student level reports are explained in detail:

- **Cover Page and Introduction** – The first page of the report provides some introductory comments from the Commissioner of Education as well as the school and district name and a table of contents. The second page gives a brief overview of the assessment system and is a good starting point for teachers new to Kentucky or anyone unfamiliar with testing in Kentucky.
- **Accountability Cycle 2002** – This page provides all the summary information pertaining to a school's accountability classification, including the growth chart unique to each school. The growth chart includes a Goal Line represented by a straight line that begins in 2000 at the baseline and ends in 2014 at 100.
- **Accountability Trend** – This page provides more detailed summary information relative to a school's accountability calculations for each year of the cycle, including academic indices for each content area, national norm-referenced test indices, non-academic indicators and the number of accountability students.
- **Disaggregation Gap Trends** – One to two pages that summarize scale score differences between certain student groups across multiple years of the assessment. A test of statistical significance is given for each comparison for each year (denoted by SD*). The number of students contributing to the calculation of each significant test is also reported.
- **Content Area Index Trends** – One page that gives comparisons/trends across multiple years *within* each content area and the overall academic index. Horizontal bar charts are used in this presentation of the data and a separate page is provided for each level (i.e., elementary, middle and high school) if necessary.
- **Academic Index Comparisons** – One page that gives comparisons of school, district, region and state academic indices for each content area and the overall academic index. Horizontal bar charts are used in this presentation of the data and a separate page is provided for each level (i.e., elementary, middle and high school) if necessary.
- **Trend Data, Number and Percent** – This page begins the "cluster" of reports for each content area. For a content area (e.g., reading), a single page gives horizontal bar charts for across-year comparisons of the percentage of students achieving Distinguished, Proficient, Apprentice (high, medium and low) and Novice (high, medium and non-performance).

- **Sub-Domain** – This is the second page of the "cluster" of reports for each content area. For a content area (e.g., reading), the school and state means for groups of items that measure each sub-domain are presented numerically and graphically. Mean item scores are calculated using both the open-response and multiple-choice questions together and are on the 0 to 4 open-response scale. A measure of standard error is provided in the graph.
- **Core Content** - The third page of the "cluster" of reports for each content area provides further detail on the performance of students by content area sub-domain and section for both multiple-choice and open-response questions. The same core content codes published in Kentucky's Core Content for Assessment are used on this report.
- **Questionnaire Data** – The fourth page of the "cluster" of reports for each content area provides student questionnaire data relevant to the content area. All questionnaire information is based on students who actually answered the questionnaire and may not represent all students who took the test.
- **Disaggregation, Performance Level Percents** – The fifth page of the "cluster" of reports for each content area provides stacked bar charts presenting a side-by-side comparison of the percentage of students achieving Distinguished, Proficient, Apprentice and Novice for a number of important student groups.
- **Mean Scale Scores/Standard Deviations** – The sixth page of the "cluster" of reports for each content area provides descriptive statistics for scale scores. Scale score means and standard deviations (presented graphically as an interval) are given for a number of important student groups.
- **Scale Score Data Disaggregation** – On the seventh page of the "cluster" of reports for each content area, scale score comparisons are provided for a number of important student groups. A standard error accompanies each scale score. In addition, differences are calculated between certain student groups (e.g., male vs. female, White vs. African-American) and a test of statistical significance is given for each comparison.
- **National Norm-Referenced Test (NRT)** – This page provides the percentage of students assigned to each accountability weight (i.e., 0, 60, 100, 140) for the National Percentile ranges 1-24, 25-49, 50-74, and 75-99, respectively.
- **NRT Data Disaggregation** – For the state mandated components of the CTBS/5 Survey, important comparisons are provided for the same student groups given on other pages of the KPR.
- **Individual Student Report** – This report informs students and parents about individual student performance on the CATS assessments.
- **Student Listing** – Yellow paper - summarizes the information included in the Individual Student Reports.
- **Item Level Report** – Blue paper - provides detailed information about student responses to individual questions on the Kentucky Core Content Test.

In later sections of this document, an image of the above reports is provided and each report is described in detail. However, before proceeding to these sections, an introduction to CATS and a review of several of the key components of CATS is given. These components include CATS Performance Levels, the Kentucky Core Content Tests (KCCT), the Accountability Index and the Long-Term Accountability System.

Introduction

In 1989 the Kentucky Supreme Court deemed the entire system of public elementary and secondary education in Kentucky unconstitutional. The Court also directed the Kentucky General Assembly to create and enact into law a new system of education that was not only constitutional but also based upon efficiency as defined by adequacy and equity. The result was House Bill 940, the Kentucky Education Reform Act (KERA), which was enacted to provide an "adequate education for all students" as mandated by the courts. One of the most comprehensive, statewide restructuring efforts ever attempted in the United States, the reform called for systemic change in finance, governance, curriculum and assessment. With regard to Kentucky's assessment system, KERA required the establishment of learning goals and identified procedures for defining and assessing the new goals. The following bullets provide an overview of the events that lead to KERA:

- November 1985 – The Council for Better Education, a nonprofit corporation formed by 66 school districts, seven boards of education, and 22 public school children sued the state of Kentucky for not providing an efficient system of education.

- October 1988 – Franklin County Circuit Court Judge Ray Corns found for the plaintiffs.

- February 1989 – Through his own actions, Governor Wallace Wilkinson issued an executive order creating a twelve-member Council on School Performance Standards. The Council was charged with determining what all students should know and be able to do and how learning should be assessed.

- June 1989 – the Kentucky Supreme Court directed the General Assembly to recreate and reestablish a "new efficient system of common schools" that complied with the Kentucky Constitution. The Court defined an efficient system of common schools as an organization that provides a "free and adequate education to all students throughout the state regardless of geographical location or local fiscal resources."

- September 1989 – the Council on School Performance Standards produced the report *Preparing Kentucky Youth for the Next Century: What Students Should Know and Be Able To Do and How Learning Should Be Assessed* and presented it to the Curriculum Committee of the Legislative Task Force charged with creating Kentucky's new system. Six broad learning goals for all students were recommended with particular emphasis on what they should be able to do. In addition, the Council recommended that the state launch a major effort to assess student performance beyond what can be measured by paper-and-pencil tests. It also was recommended that the state initiate long-range development efforts that support school reform in implementing the new learning goals.

- In 1990, the Council's recommendations were incorporated into House Bill 940, the Kentucky Education Reform Act, as a first step in redefining the school curriculum and providing what the courts required as an adequate education for all students.

- April 11, 1990 – House Bill 940 was signed by Governor Wallace Wilkinson and became law on July 13, 1990. With KERA, the General Assembly established the framework for a major revision of Kentucky's educational system. KERA required the establishment of learning goals for the educational system, provided a procedure by which those goals would be defined and assessed, and created a series of rewards and assistance to be associated with the performance of schools on those assessments.

The six learning goals established by KERA for schools within the Commonwealth are presented in the following table.

Table 1-1
Kentucky School Goals

Goal 1	Expect a high level of achievement of all students.
Goal 2	Develop student's abilities in six cognitive areas.
Goal 3	Increase school attendance rates.
Goal 4	Reduce dropout and retention rates.
Goal 5	Reduce physical and mental health barriers to learning.
Goal 6	Increase the proportion of students who make a successful transition to work, postsecondary education, and the military.

Through a two-year period of public input and review, 75 valued outcomes or performance goals were produced. The Kentucky Board of Education (KBE) approved these in December of 1991. Concerns arose about the measurability of learner goals three and four (see Table 1-1), and complaints were made about the obscurity of the wording of the valued outcomes. These concerns led to the revision and reduction of the valued outcomes to 57 in number. These were presented to the Kentucky Board of Education on May 3-4, 1994. Since that time, they have been known as the Academic Expectations. In addition to the Learning Goals and Academic Expectations, in 1992 the Kentucky Instructional Results Information System (KIRIS) was developed to measure progress toward the goals, primarily the expectations reflected in the first two goals of the act, and the non-cognitive goals outlined in goals three, four and six.

In 1998, House Bill 53 made adjustments to Kentucky's assessment and accountability programs, creating a new system call the Commonwealth Accountability and Testing System, or CATS. More specifically, an important part of this legislation directed the Kentucky Board of Education to redesign the assessment and accountability system. Through a broad and collaborative process involving educators and citizens of Kentucky, many changes were made in this new system first administered in the spring of 1999. The changes were made in order to improve the reliability and validity of the test, reduce testing time and make the system fairer and easier to understand. Those changes include, but are not limited to:

- Distributing the test components for the high school from primarily the junior year to across three grade levels;

- Reducing the contents of the Writing Portfolio in each accountability year;
- Limiting student answers on the open response to the space provided—one 8 ½" x 11" sheet;
- Including multiple-choice questions on the Kentucky Core Content Tests and weighting them 33% of the score, and weighting the open response at 67% of the Kentucky Core Content Test component of CATS;
- Giving schools incremental credit for Novice and Apprentice growth in reading, math, science and social studies; and,
- Reducing the testing window from 3 weeks to 2 weeks.

House Bill 53 shaped Kentucky's assessment and accountability system through several provisions that outline general features of a system of testing and biennial school accountability, leaving many details of implementation to various committees that were enacted by the bill. For example, the School Curriculum, Assessment, and Accountability Council (SCAAC) was created by House Bill 53 to study, review, and make recommendations concerning Kentucky's system of setting academic standards, assessing learning, holding schools accountable for learning, and assisting schools to improve their performance. The council advises the Kentucky Board of Education (KBE) and the Legislative Research Commission (LRC) on issues related to the development and communication of the Academic Expectations and Core Content for Assessment, and the development and implementation of the statewide assessment and accountability program, including the distribution of rewards and imposition of sanctions. SCAAC is composed of 17 voting members appointed by the Governor. The appointments are made to assure broad geographical representation and representation of elementary, middle, and secondary school levels, as well as equal representation of the two sexes, inasmuch as possible, and to assure that appointments reflect the minority racial composition of the Commonwealth.

House Bill 53 also required the Legislative Research Commission to appoint a National Technical Advisory Panel on Assessment and Accountability (NTAPAA), which must be composed of no fewer than three professionals with a variety of expertise in education testing and measurement. The panel advises LRC, and upon approval of the Director of the Commission, the Kentucky Board of Education and the Department of Education.

In addition to the above legislation, state law also requires KBE to set policy and promulgate regulations to implement both the assessment and accountability systems. The following are a few of the more important regulations promulgated by KBE:

703 KAR 5:010 Writing portfolio procedures.

703 KAR 5:020 The formula for determining school performance classifications and school rewards.

703 KAR 5:040 Statewide Assessment and Accountability Program; relating accountability index to school classification.

703 KAR 5:050 Statewide Assessment and Accountability Program; school building appeal of performance judgments.

703 KAR 5:070 Procedures for the inclusion of special populations in the state-required assessment and accountability programs.

703 KAR 5:080 Administration Code for Kentucky's Educational Assessment Program.

703 KAR 5:120 Assistance for schools; guidelines for scholastic audit.

703 KAR 5:130 School district accountability.

703 KAR 5:140 Requirements for school and district report cards.

Performance Levels

It can be argued that the heart and soul of CATS is the four performance levels used to describe the quality of student work. The levels, from lowest to highest, are Novice, Apprentice, Proficient and Distinguished (NAPD). In addition, the first two levels of performance in reading, mathematics, science and social studies have each been subdivided into three levels (Novice non-performance, Novice medium, Novice high, Apprentice low, Apprentice medium and Apprentice high) to better represent student performance. Kentucky law states that all schools shall expect "a high level of achievement of all students." That high level, defined by the Kentucky Board of Education, is the Proficient level.

On June 5, 2001, the Kentucky Board of Education adopted new standards for CATS. The new standards will be fully implemented this year during the 2002 CATS Accountability Cycle. While an outline of the standard setting process is provided here, a detailed Standard Setting Technical Report is available from the Kentucky Department of Education upon request.

The approximately 1600 Kentucky teachers who helped develop the standards participated in three different methods to determine the most appropriate performance standards in each of six content areas. This broad, collaborative advisory process involved teachers from every part of the state. The process itself was designed and overseen by the National Technical Advisory Panel on Assessment and Accountability, NTAPAA. The purpose was to produce a set of clear, consistent, agreed-upon recommendations for standards establishing high expectations for student achievement. As noted, this process used three different standard setting procedures and had the following six steps:

- Development of Draft Performance Descriptors
- Procedure 1 - Contrasting Groups which focused on *students' classroom performance*
- Procedure 2 - Jaeger-Mills which focused on *student work on the KCCT*
- Procedure 3 - CTB Bookmark which focused on *KCCT test items*
- Synthesis step
- Kentucky Board of Education adoption of the teacher recommended standards.

Step 1 was accomplished in two separate meetings, one in December of 1999 and the other in January of 2000. During these meetings, 88 Kentucky teachers convened to develop a set of Draft Performance Descriptors for each content area and grade level assessed by the KCCT. These Draft Performance Descriptors were developed to establish a common beginning for each of the three standard setting methods. In addition, they were developed to provide a common view of Proficient to allow for the synthesis of the three procedures, or more specifically, the synthesis of the three sets of cut-score recommendations resulting from the three procedures. Perhaps more importantly, the Draft Performance Descriptors were developed with the end product in mind, that is, *to assist teachers in aligning instruction with assessment expectations*. Along these lines, the Draft Performance Descriptors, now called Performance Descriptions, were refined during standard setting (as part of the procedures) to assure congruence between the demands for students as seen in the content/cognitive descriptions and the demands of the actual assessment. These descriptions by grade level and content area can be found on the Kentucky Department of Education's (KDE) website at http://www.kde.state.ky.us/.

Step 2, the Contrasting Groups procedure, took place in April 2000 and involved 951 teachers. Using the same draft descriptors developed in Step 1, participants used the descriptors to evaluate their own students' classroom performance. Student performance on homework assignments, teacher made tests, classroom participation, etc., was evaluated using the draft descriptors. In other words, these teachers used their own professional judgment and the draft descriptors to categorize their students as Novice, Apprentice, Proficient or Distinguished. If the decision to place a student into one of these four categories was too difficult, teachers were allowed to place the student in one of three borderline categories, i.e., Novice/Apprentice, Apprentice/Proficient or Proficient/Distinguished. While the other two procedures involved teachers coming together in a face-to-face meeting (see below), the Contrasting Groups did not. That is, no "formal" training for participants occurred as did in the other procedures. In addition, while teachers were provided with written directions on how to apply the Draft Descriptors for making their judgments about students, it is possible that eight years of experience with the old KIRIS cut-scores may have contributed to the judgment of teachers.

Step 3, the Jaeger-Mills procedure, took place in October 2000 and involved 312 teachers who came together for a three-day meeting. The main focus for these teachers was actual complete student work in a content area from the Spring 2000 administration of the KCCT. These teachers also used the Draft Descriptors to categorize student work. Teachers categorized 60 sets of complete student work, each set containing responses to 6 open-response questions and 24 multiple-choice questions. Using the Draft Descriptors, teachers systematically placed each set of student work into one of 12 categories, a low, middle and high category for each of the four performance levels (NAPD). Cut-points for the Jaeger-Mills procedure were obtained by calculating the median value for the "high" and "low" categories of adjacent performance levels, and then taking the middle point between these two values. While the Jaeger-Mills procedure worked quite well, more training time would have been desirable. Similarly, more time refining the descriptors would have also been useful. Finally, in some content areas, the assessment may not have allowed students to demonstrate Distinguished performance relative to the draft descriptors. For example, it is difficult for a single item, or even a set of items, to adequately assess the integration of concepts across content areas or to assess the actual use of manipulatives (e.g., equipment used in science or maps for social studies). This latter observation was very important and led to further refinement of the descriptors to assure congruence between the descriptors and the assessment.

Step 4, the CTB Bookmark procedure, took place in December 2000 and involved 290 teachers who came together for a two-day meeting. The main focus for these teachers was KCCT test items from the Spring 2000 assessment. Prior to the meeting, for each grade level and content area, a book of items was compiled so that the items were ordered by difficulty based on how well students performed on the items in Spring 2000. Items that were easy for students appeared early in the book, while items that were more difficult for students appeared later in the book. Each of the booklets contained both open-response and multiple-choice items. Once again, teachers used the Draft Descriptors as a starting point. The task for each teacher was to literally place a "bookmark" within the book to indicate the location where a correct response to a particular question would, in the teacher's judgment, place a student into the next higher performance category. Each teacher placed three bookmarks within a book, one for each cut-point, or put another way, one to denote the transition from Novice to Apprentice, from Apprentice to Proficient and from Proficient to Distinguished. Because in Item Response Theory both test items and test takers are put onto the same numerical scale (i.e., the scale score scale), the three bookmarks placed by each teacher translated into three cut-points. Calculating the median value across the teachers within a grade level and content area provided the cut-points from the CTB Bookmark procedure. Two final points about the CTB bookmark procedure are that teachers were given the opportunity to discuss their recommendations prior to submitting final cut-point values and teachers may have been limited by the fact that only part of the total item pool was available for use in the procedure (only 1/3 of the total assessment item pool could be used to construct the ordered item booklets).

Step 5, the Synthesis step, took place in February 2001 and involved 132 teachers who came together for a three-day meeting. For a teacher to participate in the Step 5 Synthesis, the teacher had to have already participated in one of the previous three procedures. The Synthesis step achieved many important objectives. These objectives are summarized in the following bullets where participants had to:

- Understand what had been accomplished in the first four steps of the standard-setting process.
- Evaluate and discuss the instructional implications of the three standard-setting methods.
- Study the recommended cut-scores within the context of impact data.
- Make a subject/grade-level recommendation for the appropriate cut-scores.
- Discuss recommended cut-scores with other subject areas within the same grade level.
- Discuss recommended cut-scores with other grade levels within subject areas.
- Make a final recommendation with impact data to the Kentucky Board of Education.
- Summarize the instructional implications of the cut-scores, and refine the descriptors to fit the cut-score.

The above standard setting project, which took over 18 months to complete, was unique in that it used three different methods to determine the standards. While in retrospect there were some limitations in each method, all three methods were well implemented and consistent with the design as established by the state's National Technical Advisory Panel for Assessment and Accountability. The data from all three methods were valuable in establishing the final recommendations forwarded to the Kentucky Board of Education. In addition to the specific standard setting steps outlined above, between May 10 and May 28, 2001, more than 3,000 people—2,891 identifying themselves as educators—responded to a Kentucky Department of Education online survey about the standards setting process. Slightly more than 32 percent of the respondents said they were "very comfortable" or "comfortable" with the standards setting process. Only 16 percent said they were uncomfortable with the process. A total of 3,184 people commented on the process by which the standards were developed and/or reviewed the descriptions and submitted comments for the Kentucky Board of Education. The Board in reviewing the standards considered this input. On June 5, 2001, as the final step in the standard setting process (Step 6), the Kentucky Board of Education adopted the new teacher recommended standards.

As a final note, one of the more important products, if not the most important product, generated from the standard setting process was a set of Instructional Summaries. In fact, in the Synthesis step, *three* sets of Draft Instructional Summaries were provided to teachers, each set based upon the cut-points derived from one of the three procedures (Contrasting Groups, Jaeger-Mills, and CTB Bookmark). *Using the different sets of Draft Instructional Summaries allowed Step 5 participants to evaluate cut-scores without looking at any other data* (e.g., scale scores, distributions of student scores, etc.). It was not until the final day of the Synthesis step meeting that teachers were allowed to view and discuss actual numbers. The following bullets summarize the most important considerations regarding the Draft Instructional Summaries:

- Were improved upon by teachers during the standard setting process.
- Reflect NAPD performance standards resulting from each of the standards setting methods.
- Gave the Synthesis step a beginning point.
- Content – Using the cut-scores identified by each method, an effort was made to summarize the content of items that located or fell within in each performance level (NAPD).
- Cognitive – Using the cut-scores identified by each method, an effort was made to summarize the cognitive skills associated with each performance level (NAPD).

In conclusion, the new standards are important because they define what Novice, Apprentice, Proficient and Distinguished levels of performance mean. They clarify for teachers, students and parents how the Kentucky Core Content Test evaluates student work, and they explain for students what is expected of them. The final cut scores, for each grade and content area, are in Appendix A. The Kentucky scale ranges from 325 to 800 in all grades and content areas. Each scale was set to have a mean of approximately 500, and standard deviation of approximately 50 in 1999. The mean and standard deviations varied some from grade to grade because of relationships to previous KIRIS scaling. The State Board adopted descriptions of Novice, Apprentice, Proficient and Distinguished by grade level and content area can be found on the Kentucky Department of Education's (KDE) website at http://www.kde.state.ky.us/.

Measures and Indicators

Both academic content-based and non-academic measures are used in CATS. These measures include custom, criterion-referenced tests in reading, mathematics, science, social studies, arts and humanities, practical living/vocational studies and writing. Non-academic measures include attendance rate, retention rate, dropout rate and transition to adulthood. (Note that transition to adulthood data is collected in the fall of each year via a short survey completed by school personnel. Measures include the number of graduates planning to enter college, the military, or an alternative vocation.) The above multiple measures were selected to provide as complete a snapshot of schools as possible and to communicate to schools the importance of each measure and indicator in terms of resources and instructional programs.

Writing Portfolio

As part of the assessment, students developed portfolios in writing. The "holistic" performance level scores submitted by teachers trained to evaluate portfolios are presented on the Individual Student Report, Student Listing and the Kentucky Performance Report. Please note that information on the instructional analysis for each student's writing portfolio was not collected in this year's assessment, and therefore is not reported. Portfolios support teachers' efforts to actively engage their students in performance-oriented educational activities. Therefore, the Department considers the effective implementation of the portfolio assessment to be a high priority.

During the summer of 2002, the Kentucky Department of Education conducted a Writing Portfolio Audit at grades 4, 7 and 12. One hundred and one (101) schools throughout the state were selected to participate in the audit. Participating schools and their students are identified by a statement on the Individual Student Report and at the end of the Student Listing. The Writing Portfolio scores on these reports are the scores determined by the audit. Scores for the schools not participating in the audit are the scores assigned by teachers.

Accommodations and Modifications

Kentucky's assessment program offers accommodated or modified assessments for students who qualify. The accommodation/modification must be stipulated in the student's Individual Education Plan (IEP) or 504 and must have been used with the student throughout the school year. For example, if a student's IEP allows a scribe during regular instruction, the student will be allowed to have a scribe for the statewide assessment. Other accommodations or modifications, when consistent with the normal on-going delivery of instruction, may include:

- Reading text in English
- Paraphrasing directions for tasks in English
- Oral word-for-word translation of text
- Administering assessments in small groups
- Use of foreign language dictionaries
- Use of word processor or typewriter
- Use of grammar or spell-checker.

In addition to the above accommodations or modifications, in 2002 Kentucky had a two-year exemption for students whose primary language was not English. More specifically, Limited English Proficient (LEP) students must have been in an English-speaking school for two full years preceding the year of the assessment before participating in the assessment with or without accommodations or modifications. Because this policy is not in alignment with federal regulation (i.e., Title I and IDEA), Kentucky applied for and was granted a one-year exemption while the state develops policies for serving and assessing LEP students. Depending upon the current reauthorization, the state plans on allowing only a one-year exemption for LEP students prior to participating in the 2003 statewide assessment.

Alternate Portfolio

Students who cannot participate in the regular assessment, even with accommodations, are required to submit an alternate portfolio. These students usually have profound cognitive disabilities and the alternate portfolio is the only way they can participate in the assessment and accountability systems. With few exceptions, all students in Kentucky must participate in the regular assessment or the alternate portfolio. Only a small number of students qualify each year for an exemption from testing.

Testing Exemptions

Students can receive a medical exemption if certain criteria are met (e.g., the stated medical condition *cannot* be the student's disability) and a physician determines that the student cannot physically take the test or that participation would be harmful to the child. Foreign exchange students are also exempt from the statewide assessment. All together, less than one percent of students statewide are exempted each year from Kentucky's assessment program.

Spring Testing

All testing is completed in the spring of each year, including the administration of a norm-referenced test (CTBS/5 Survey Edition) in grades 3 (end of primary), 6 and 9. Beginning with the 2002 Accountability Cycle, the results of the norm-referenced test contributed to the calculation of a schools accountability index. Recall that the long-term goal for every school in the state is Proficiency as defined by the Kentucky Board of Education. This goal of Proficiency translates into a school accountability index value of 100 (i.e., the goal for the state is for each school to achieve an accountability index of 100 by 2014). Each of the measures/indicators mentioned above are combined into a composite to obtain a school's accountability index (see Kentucky's Accountability Index section below).

Kentucky Core Content Test

The measurement that contributes most to the calculation of a school's accountability index is the Kentucky Core Content Test (KCCT). The table on the following page summarizes the grades and content areas tested by the Kentucky Core Content Test, including the number of open-response and multiple-choice questions asked on each of six (6) forms of the KCCT (12 forms each for arts and humanities and practical living/vocational studies). At all grade levels where reading, mathematics, science and social studies are tested, seven open-response and twenty-eight multiple-choice questions are given to each student (one open-response and four multiple-choice questions are pre-test questions and are not included in student scores or school accountability calculations). At the grade levels where arts and humanities and practical living/vocational studies are administered, three open-response and twelve multiple-choice questions are given to each student (one open-response and four multiple-choice questions are pretest questions and are not included in student scores or school accountability calculations). Because there are six forms of the test and the forms generally do not overlap, this means that for accountability purposes there are 36 open-response items and 144 multiple-choice items administered per grade level/content area for reading, mathematics, science and social studies. For arts and humanities and practical living/vocational studies, there are 24 open-response items and 96 multiple-choice items administered per grade level/content area because there are 12 non-overlapping forms of the test. Note that multiple-choice scores in each content area are included in school accountability calculations. Finally, students at grades 4, 7 and 12 select and respond to one of two on-demand writing prompts offered during the test.

2001-2002 ASSESSMENT COMPONENTS									
Grade	**Kentucky Core Content Test**							**Portfolio**	
	Rdg	**Math**	**Sci**	**Soc St**	**Wrtg**	**A&H**	**PL/VS**	**Wrtg**	**Alt***
4	6 OR* 24 MC		6 OR 24 MC		X*			X	X
5		6 OR 24 MC		6 OR 24 MC		2 OR 8 MC	2 OR 8 MC		
7	6 OR 24 MC		6 OR 24 MC		X			X	
8		6 OR 24 MC		6 OR 24 MC		2 OR 8 MC	2 OR 8 MC		X
10	6 OR 24 MC						2 OR 8 MC		
11		6 OR 24 MC	6 OR 24 MC	6 OR 24 MC		2 OR 8 MC			
12					X			X	X

*** OR denotes Open Response, MC denotes Multiple Choice; "X" denotes that On-Demand Writing (or the Writing Portfolio) was administered; "Alt" denotes participation in the Alternative Portfolio program.**

Open-response items are scored on a 0 to 4 scale for each item. For example, an off-topic response to an open-response item would receive a 0. Students must respond with some relevant information that is above and beyond merely restating the question to receive a score above 0. An outstanding response to an open-response item, one that is correct, thorough and well communicated, would receive a higher score, perhaps a 3 or a 4. Each open-response item has its own unique scoring rubric. The Department's scoring contractor trains professional scorers to score all the open-response items on the KCCT. It takes over 800 scorers more than two months to score the tens of thousands of student responses obtained each year from the administration of the KCCT.

In Kentucky, open-response items are very important to the statewide assessment because they model the type of instructional strategies the state would like to see in Kentucky classrooms. While students who score mostly 3s and 4s on the open-response items within a content area have a higher probably of scoring a Proficient or Distinguished within that content area, *the item scores of 1, 2, 3 and 4 DO NOT correspond to Novice, Apprentice, Proficient and Distinguished (N, A, P and D), respectively*. Recall from the standard setting discussion above that cut-scores for N, A, P and D were obtained from teacher's judgments of the totality of a student's work, or from reviewing numerous test items provided in sequential order. A score of 4 on one item in a KCCT content area does not lead to a Distinguished performance level by itself.

The KCCT also has multiple-choice items that are scored as correct or incorrect. Multiple-choice items were added to the KCCT to increase content domain coverage and to increase the reliability of scores within a content area. The same Kentucky teachers (Content Advisory Committee) that develop the open-response items for the KCCT also develop the multiple-choice items. In fact, the same item-development procedures are followed for both types of item formats. For example, the same rules for strict adherence to the *Core Content for Assessment* are followed, as well as the item selection parameters relating to item difficulty. Because of this, the multiple-choice items on the KCCT have different characteristics than the multiple-choice items on a nationally norm-referenced test such as the CTBS5/Survey. KCCT multiple-choice items match Kentucky's core content much better and the items are generally more difficult for students than the items on a nationally norm-referenced test.

Thus far, the only KCCT characteristics mentioned have related to test format (e.g., the KCCT has multiple forms) and item type (open-response vs. multiple-choice items). In addition, the only scoring mentioned thus far relates to simple item raw scores. According to a scoring rubric, a student can get a raw score of 0, 1, 2, 3 or 4 on an open-response item and a 1 or a 0 (i.e., correct or incorrect) on a multiple-choice item. In the next section the discussion centers on how you can actually go from simple raw scores to an accountability index that summarizes a schools progress toward the state's goal of Proficiency.

Kentucky's Accountability Index

The long-term goal for every school in the state is Proficiency as defined by the Kentucky Board of Education. The goal of Proficiency translates into a school accountability index value of 100. More specifically, the goal for the state is for each school to achieve an accountability index of at least 100 by 2014. In the Long-Term Accountability Model discussed in a later section, intermediate targets that will eventually take a school to the goal of 100 are set biennially, or every two years starting in 2002. As such, there are seven biennia or accountability cycles between 2002 and 2014 (i.e., 2002, 2004, 2006, 2008, 2010, 2012 and 2014). The major characteristics of the accountability model is that it involves (a) an index, (b) comparisons or a measure of growth between successive groups, (c) criteria that are applicable to the whole school and (d) differential weighting of indicators.

With respect to the Long-Term Accountability Model, the previously discussed indicators are combined to create an accountability index that is unique to each school. The progression of how this happens begins with simple number-correct raw scores and ends with an accountability index that summarizes a school's progress toward the state's goal of Proficiency. To state this progression in one sentence, raw scores give rise to scale scores, scale scores have been related to Novice, Apprentice, Proficient and Distinguished (NAPD) performance levels (via standard setting and cut-scores), NAPD's get weighted numerically and combined within each content area, and finally, the content areas are weighted and combined to form a school's accountability index. This progression is summarized below:

Raw Scores → Scale Score → Cut Scores/NAPD → Numerical Weights for NAPDs → Indices

The following 4 steps describe this process in more detail.

Step 1 - Raw Scores Give Rise to Scale Scores

Raw scores are the simplest scores to understand because they have the most direct connection to the actual questions on a test. Test questions are either right or wrong, or in the case of open-response questions, there is the sequence of increasingly better answers worth from 1 to 4 raw score points. These are the same types of item raw scores that teachers commonly use in their classrooms. Similarly, teachers add up all the correct responses for each student, which results in a number correct raw score that summarizes the overall performance of each student on the test. The KCCT also adds up all the correct responses within a content area for each student and provides a number correct raw score that summarizes the student's performance. For example, for the content areas of reading, mathematics, science and social studies:

- 6 open-response items (each scored 0-4) gives a possible raw score range of 0 to 24
- 24 multiple-choice items (each scored 0-1) gives a possible raw score range of 0 to 24.

Say Student 1 scores:

- 17 open-response points (out of 24) and 16 multiple-choice items correct (out of 24). Or, more specifically, for Student 1:

- 17 open-response points (out of 24) is weighted double, so 17 X 2 = 34. 16 multiple-choice items correct (out of 24) is weighted only once, so 16 X 1 = 16.

Add 34 and 16 together (i.e., 34 + 16 = 50) and you have Student 1's raw score.
For reading, mathematics, science and social studies the possible raw score range goes from 0 to 72 because open-response items are weighted double in CATS. (Recall that open-response items model the type of instructional strategies the state would like to see in Kentucky classrooms.) As such:

- Open-response items can equal up to 48 raw score points whereas
- Multiple-choice items can equal up to 24 raw score points
- 48 + 24 = 72 possible raw score points.

Similarly, for the content areas of arts and humanities and practical living/vocational studies:

- 2 open-response items (each scored 0-4) gives a possible raw score range of 0 to 8
- 8 multiple-choice items (each scored 0-1) gives a possible raw score range of 0 to 8.

Say Student 2 scores:

- 6 open-response points (out of 8) and 7 multiple-choice items correct (out of 8). Then, for Student 2:

- 6 open-response points (out of 8) is weighted double, so 6 X 2 = 12. 7 multiple-choice items correct (out of 8) is weighted only once, so 7 X 1 = 7.

Add 12 and 7 together (i.e., 12 + 7 = 19) and you have Student 2's raw score.

For arts and humanities and practical living/vocational studies the possible raw score range goes from 0 to 24 because open-response items are weighted double in CATS. As such:

- Open-response items can equal up to 16 raw score points whereas
- Multiple-choice items can equal up to 8 raw score points.
- 16 + 8 = 24 possible raw score points for these *two* content areas.

Wouldn't it be nice to have only one form of the KCCT, and then everything could be done in raw score units. If students took only one form of the test, there would be no reason to use anything but number correct raw score. Unfortunately, there are many good reasons for why it is not possible to administer only one form of the KCCT. First and foremost, to obtain the content coverage necessary for a fair high-stakes assessment and accountability system, a single form format would require too much testing for any one student, especially across multiple content areas. This is truer for Kentucky's program than for any other state program because the foundation of the assessment is our open-response questions. Can you imagine a student taking 36 open-response questions in a single content area! This is the number of open-response questions each student would have to take per content area because each form of the test has 6 open-response questions and there are six forms of the test (i.e., 6 X 6 = 36). In addition, serious test security issues arise when only one form of a test is administered statewide. For example, students coping off each other can become a problem (they are all looking at the same test items) and security across schools and school districts becomes more difficult. Some large-scale testing programs that have only one form during an administration limit their testing window to only one day or even only one morning or afternoon (e.g., SAT, ACT and AP exams).

Because it is necessary to have multiple forms of the KCCT, the question then becomes, how do I know one form of the test isn't more or less fair than another form? Also, how do you combine all of the different forms administered into one thing that makes sense? The answer to these questions is Item Response Theory. The use of Item Response Theory (IRT) is by no means unique to Kentucky. IRT was invented long before education reform in Kentucky. In fact, Kentucky's use of this technology is a very standard use. The following example demonstrates why scale scores are so important for "leveling the playing field" on the Kentucky Core Content Test:

Student	Raw Score	Form	Score
1	50	Form 1	586
2	50	Form 6	583
3	50	Form 1	586
4	69	Form 1	691
5	65	Form 2	657
6	38	Form 4	536
7	39	Form 3	536
8	70	Form 3	680

Provided above are the number-correct raw scores and accompanying scale scores for eight students who each took one of the six forms in a content area of the KCCT. Inspection of this data reveals several observations. First, the same raw score on a different form can, and usually will, generate a different scale score. Raw scores are converted to scale scores to address the minor differences in difficulty among the six test forms. So while students 1 and 2 each obtained a raw score of 50, student 1 received a few more scale score points than student 2 (i.e., 586 vs. 583) because Form 1 was slightly more difficult than Form 6 at this particular point of the scale score range. Note how this did not put student 2 (the student that took Form 6) at a disadvantage because the student had an equal opportunity to score a 583 on any form of the test. Had the student taken Form 1, because this Form is slightly more difficult than Form 6, the student probably would have scored a few raw score points lower than 50.

Similar to the above observation, note the difference between students 6 and 7. These two students received the same scale score (i.e., 536) but different raw scores. Student 7 received a raw score of 39, one point higher than student 6 who received a raw score of 38. Had student 7 taken Form 4, Item Response Theory would predict that this student would receive a raw score of 38 (one point less than with Form 3) because this student's ability in scale score units is 536. The main point from these two examples is that there are minor differences in difficulty among the six form of the test, but *scale scores produced on different Forms mean the same thing.* Two students who receive the same scale score at the same grade level in the same content area are said to have the same ability level, regardless of the Form they took.

As previously stated in the Measures and Indicators section, there are multiple forms of the test for each grade level and content area assessed and the forms generally do not overlap. To compensate for small differences in difficulty among forms, and to bring all forms of a test for a grade level and content area onto the same scale, Item Response Theory is used. As such the underlying scale for the KCCT is not number-correct raw score, but rather a scale score scale that ranges from approximately 325 to 800 with 500 being the middle of the scale.

Step 2 - Scale Scores Have Been Related to Performance Levels

It can be argued that the heart and soul of CATS is the four performance levels used to describe the quality of student work. The four levels, from lowest to highest, are Novice, Apprentice, Proficient and Distinguished or NAPD. During standard setting (see Performance Levels section above), these four performance levels were related to, or mapped onto, the range of scale scores for each grade level and content area test. In addition, beginning in 1999, the first two levels of performance in reading, mathematics, science and social studies were each subdivided into three levels (Novice non-performance, Novice medium, Novice high, Apprentice low, Apprentice medium and Apprentice high) to better represent student performance.

Step 3 - NAPD's Get Weighted Numerically and Combined

Students taking a test in a particular content area are assigned to one of the above eight performance levels. This is the official "score" that gets reported for the student. For example, a fourth grade student might receive an Apprentice in reading and a Proficient in science. For reporting in the aggregate and for accountability purposes only, the following conversion table is used for transforming NAPD's into a numerical scale that ranges from 0 to 140:

Performance Level	Weight
Novice Non-performance	0
Novice Medium	13
Novice High	26
Apprentice Low	40
Apprentice Medium	60
Apprentice High	80
Proficient	100
Distinguished	140

If the following distribution (or percentages) were obtained by fourth graders administered the reading test in a particular school, the calculations would be:

Performance Level	Weight	Percentage	Calculation
Novice Non-performance	0	5%	0 X .05
Novice Medium	13	10%	13 X .10
Novice High	26	15%	26 X .15
Apprentice Low	40	20%	40 X .20
Apprentice Medium	60	25%	60 X .25
Apprentice High	80	15%	80 X .15
Proficient	100	8%	100 X .08
Distinguished	140	2%	140 X .02
Total of Sum		100%	51.0

As demonstrated in the above table, the weights for the NAPD's are multiplied by the percentage (or rather the proportion) of students at each performance level and then simply summed across the performance levels. The resulting content area index for fourth grade reading in this school is 51.0. The same procedure is used for calculating the "academic" index for each content area. Note the direct connection between the performance levels and a content area or academic index. If every fourth grade student in the school had scored Proficient (i.e., the state goal) on the reading test, the school reading index would be 100 (or at the state goal). As seen in the next step, this connection is maintained all the way through to a school's weighted accountability index.

Step 4 - Content Areas Get Weighted and Combined

Once an academic index has been calculated for all content area tests administered within a school, the school's accountability index for a particular year can then be determined. The weights used to calculate a school's accountability index vary slightly depending upon whether the school is an elementary, middle or high school. The following formula reflects the weighting of components at the *high school* level (elementary and middle school have different weights).

Given the following definition of terms in the formula:

RD = Reading
MA = Mathematics
SC = Science
SS = Social Studies
AH = Arts & Humanities
PL = PL/VS
WR = Writing
NA = Non-academic
NRT = CTBS Survey

To calculate the index for a given year:

$$\text{Accountability Index} = .95*[(RD*.15) + (MA*.15) + (SC*.15) + (SS*.15) + (WR*.15) + (AH*.075) + (PL*.075) + (NA*.10)] + .05*(NRT)$$

The weights used for calculating an Accountability Index sum to one. In the above formula, the weights within the brackets add to one but are then multiplied by .95. The NRT component of the assessment (CTBS5/Survey) makes-up the remaining 5%. (While the combination of weights could have been multiplied out (e.g., .95 * .15 = .1425), the above formula helps to show the content area weights before the NRT is added).

At the high school level, the non-academic component (denoted NA above) is weighted 10% and is comprised of the following components with the following weights:

Non-Academic Index (10%)	
Attendance Rate	2.00%
Retention Rate	0.50%
Dropout Rate	3.75%
Successful Transition to Adult Life	3.75%

The NRT component is based upon the CTBS/5 Survey (state required components) Total Battery National Percentile. The "index" for the NRT is an average of student scores assigned as follows:

Score	National Percentile
0	1 - 24
60	25 - 49
100	50 - 74
140	75 - 99

Note that the assignment of such scores puts the NRT onto the 0 to 140 scale of the other content areas. As previously mentioned, the mean score for students on this new scale is then weighted 5%.

Long-Term Accountability Model

The above formula, or weighted composite, for the Accountability Index is for one year only. Recall that the intermediate targets which will eventually take a school to the goal of 100 are set biennially, or every two years. In other words, the above Accountability Index calculations have to be performed for both years of the baseline and both years of the subsequent target years. The Long-Term Accountability baseline index is the arithmetic mean of the Accountability Index for 1999 and for 2000, i.e., (1999 Index + 2000 Index)/2. In the same way, the growth index for the CATS Accountability Cycle ending in 2002 is the arithmetic mean of the Accountability Index for 2001 and for 2002, or (2001 Index + 2002 Index)/2. The growth index for the Accountability Cycle ending in 2004 is the arithmetic mean of the Accountability Index for 2003 and for 2004, or (2003 Index + 2004 Index)/2. The growth indices for the remaining 5 biennia or Accountability Cycles are calculated in the same way.

Remember that the long-term goal for all schools is to reach Proficiency, or a growth index of 100, by 2014. The interim targets established for each two-year Accountability Cycle beginning in 2002 and ending in 2014 represent a requirement that achievement improve by a set amount each year. Along these lines, each school has its own unique set of growth targets. Growth targets are calculated using the following formulas:

For 2002: (((100-baseline)/7) * 1) + baseline
For 2004: (((100-baseline)/7) * 2) + baseline
For 2006: (((100-baseline)/7) * 3) + baseline
For 2008: (((100-baseline)/7) * 4) + baseline
For 2010: (((100-baseline)/7) * 5) + baseline
For 2012: (((100-baseline)/7) * 6) + baseline
For 2014: (((100-baseline)/7) * 7) + baseline

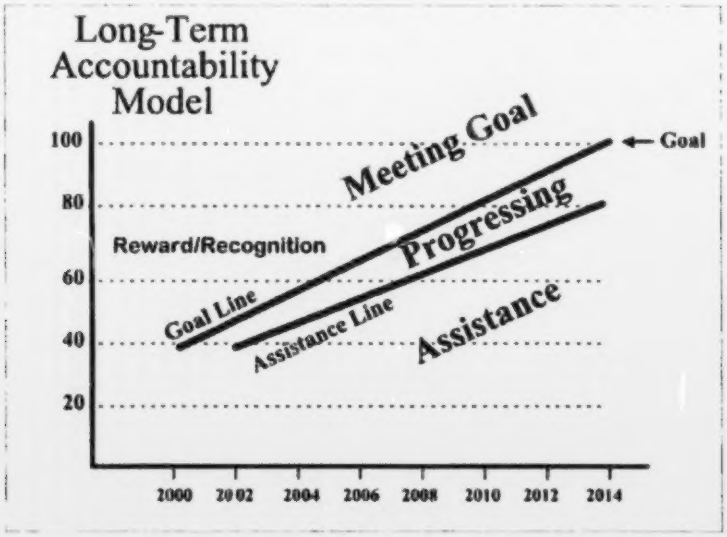
For example, given a baseline index of 51, the calculations would be:

For 2002: (((100-51)/7) * 1) + 51 = 58
For 2004: (((100-51)/7) * 2) + 51 = 65

For 2006: (((100-51)/7) * 3) + 51 = 72
For 2008: (((100-51)/7) * 4) + 51 = 79
For 2010: (((100-51)/7) * 5) + 51 = 86
For 2012: (((100-51)/7) * 6) + 51 = 93
For 2014: (((100-51)/7) * 7) + 51 = 100.

In this example, the school's growth index in Accountability Cycle 2002 would be compared to the growth target of 58. Similarly, the school's growth index in Accountability Cycle 2004 would be compared to the growth target of 65, and so on. The presentation of a school's growth targets is simplified by presenting them in the following graphic. Note that in this example, the growth targets are based upon a baseline index of 40.

Long-Term Accountability Growth Chart:



The following bullets summarize some important points about the above graphic and several other features of the Long-Term Accountability Model:

- The Goal Line represents the point above which schools become eligible for monetary rewards. Notice how it is represented by a straight line that begins in 2000 at the baseline and ends in 2014 at 100.
- The Assistance Line represents the point below which a school becomes eligible for assistance from the state. A straight line that begins in 2002 at the baseline and ends in 2014 at 80 represents this line.
- Both of the above lines (the Goal Line and the Assistance Line) have a standard error associated with the line that ranges from approximately .5 to 3.0 depending upon school level (elementary, middle and high school) and school size. (The standard error is represented by the thickness of the line.)

- Schools between the Goal Line and the Assistance Line are considered Progressing and are held harmless in the accountability system.
- For a school to be eligible for rewards, it must also meet the Novice reduction and dropout criteria. With regard to Novice reduction, schools must reduce their percent of Novices on a schedule so that by 2014, the school has 5% percent or less of its students scoring Novice. With regard to the dropout criteria, high schools must have a dropout rate less than or equal to 5.3%, or reduce their percent dropout by 0.5%, but still have a dropout rate less than or equal to 6.0%.
- The Long-Term Accountability Model also has provisions for establishing a set of one time Recognition points and also defines the requirements for being a "Pace Setter" school.

Other important considerations regarding Kentucky's Accountability Model include:

- Because many schools in Kentucky are small, two years of data are combined to form both the baseline and the growth indices. Combining two years of data addresses some of the stability issues related to estimating achievement for small schools. The Long-Term Accountability Model is used to evaluate all regular schools (and students within alternative programs) regardless of school size.
- Results from non-standard administrations of the assessment (accommodated or modified testing) are included in accountability calculations the same way as results from standard administrations of the tests.
- While K-2 schools do not participate in the assessment program which starts in grade 3 (end of primary), these schools can receive reward money if the regular or accountable school the K-2 school feeds into qualifies for rewards. (It should be noted that there were only 19 K-2 or K-3 schools in Kentucky during the 1999/2000 school year. Of those, seven K-3 schools actually had waivers in place to have their accountability scores included with the "receiving" school.)
- The four non-academic components (i.e., attendance, retention, dropout and successful transition to adult life) are not computed on the 0 to 140 scale. Rather, these components are each put onto a 0 to 100 scale. More specifically, the values for attendance and successful transition to adult life are the actual percentages reported, whereas the values entered into calculations for retention and dropout are 100 minus the actual percentage calculated. Because of the minimal weighting attributed to non-cognitive measures, the impact of this on a school's overall, weighted accountability index is slight.
- For Title I, an index is created for each district based only upon the schools within the district that receive Title I funds. This index is evaluated for purposes of federal reporting.

As a final note, results from the Alternate Portfolio, Kentucky's means of assessing the instruction provided to students with significant disabilities, are scored using the same performance levels as the content area tests (i.e., NAPD). An Alternate Portfolio is submitted only once at the elementary level, once at the middle school level, and once at the high school level. *At each of these levels, a student's performance level (N, A, P or D) weight contributes to all content areas.*

For example, if an Alternate Portfolio student receives a Proficient, for calculation purposes, it is as if the student received a Proficient (weight of 100) in all content areas of the assessment at the grade level. In this way, Alternate Portfolio students contribute the same amount to accountability as any other regular education student, although that contribution happens within one calendar year and not across several years (e.g., fourth and fifth grade or seventh and eighth grade). The main justification for this is the importance of including all students in assessment and accountability. Similarly, the scores for students who receive accommodations or modifications are treated the same as students who received no accommodations or modifications. In Kentucky, the inclusion of all students is weighed more heavily, i.e., is more important in terms of consequential validity, than the small challenge to construct validity that may result when alternate and accommodated student scores are included with all other student scores.

Explanation of Reports

This section provides detailed information on how to interpret and use the September 2002 assessment and accountability results provided by the Kentucky Department of Education (KDE). The data in these reports were constructed from information provided by many sources: students, schools, district offices, the Kentucky Department of Education and testing contractors. Many of the report pages discussed below are part of the Kentucky Performance Report (KPR). The KPR is designed to show performance for all content areas at the elementary, middle and high school levels. Therefore, most school and all district Kentucky Performance Reports will contain data from at least two different grades (e.g., grades 4 and 5 at the elementary level).

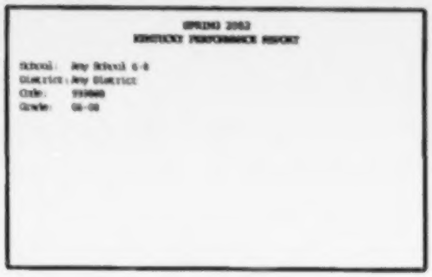
Note that school staff must review the data on the "Student Listing" report to ensure all students who tested last spring are represented accurately on the reports. If your school/district has concerns about the data, please contact KDE, Division of Assessment Implementation at 502/564-4394. The Kentucky Department of Education will explain the procedures and assist schools in correcting data to ensure accurate school Academic and Accountability Indices.

Cover Page and Introduction

The first page of the Kentucky Performance Report (KPR) provides some introductory comments from the Commissioner of Education as well as the school and district name, school code, grade range covered in the report and a table of contents. The Commissioner's statement generally includes commentary on important policies related to assessment and accountability in Kentucky. For example, the inclusion of Kentucky teachers in test development, the value of the new performance standards to instruction, and the goal of 100, or proficiency by 2014.

The second page of the KPR gives a brief overview of the assessment system and is a good starting point for teachers new to Kentucky or anyone unfamiliar with testing in Kentucky. Some of the topics introduced on this page include the content areas tested at each grade level, the number of multiple-choice and open-response questions assessed in each content area and their respective weight in school accountability, and the particular students a school is held accountable for. The first two pages of the KPR are presented on the following page in the sequence they appear in actual reports.

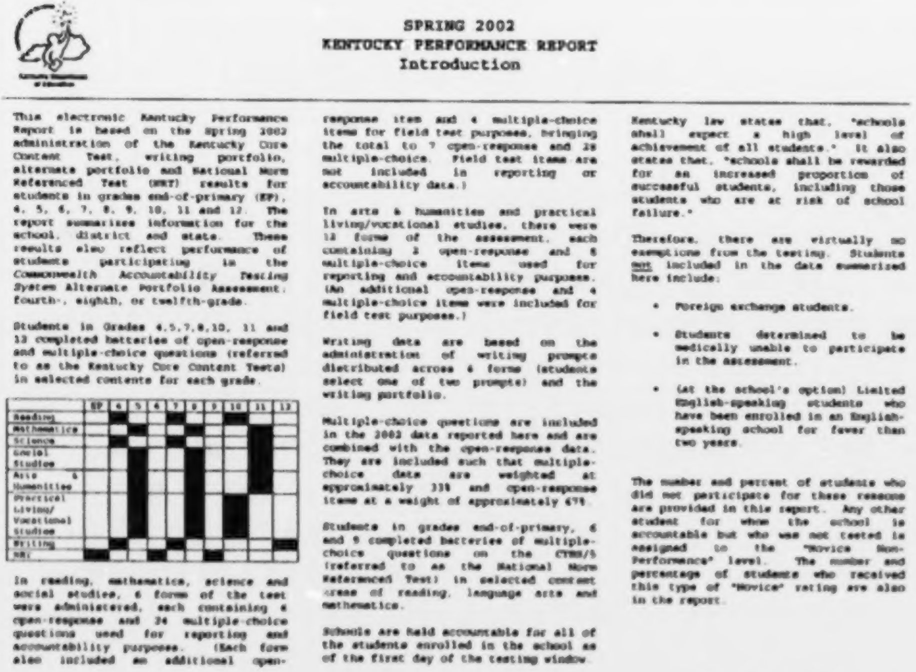
KENTUCKY DEPARTMENT OF EDUCATION
CAPITAL PLAZA TOWER - 500 MERO STREET - FRANKFORT, KENTUCKY 40601
Gene Wilhoit, Commissioner of Education



SPRING 2002
KENTUCKY PERFORMANCE REPORT

School: Any School 6-8
District: Any District
Grade: 06-08

CONTENTS OF THIS REPORT



SPRING 2002
KENTUCKY PERFORMANCE REPORT
Introduction

This electronic Kentucky Performance Report is based on the Spring 2002 administration of the Kentucky Core Content Test, writing portfolio, alternate portfolio and National Norm Referenced Test (NRT) results for students in grades end-of-primary (EP), 4, 5, 6, 7, 8, 9, 10, 11 and 12. The report summarizes information for the school, district and state. These results also reflect performance of students participating in the *Commonwealth Accountability Testing System* Alternate Portfolio Assessment, fourth-, eighth, or twelfth-grade.

Students in Grades 4,5,7,8,10, 11 and 12 completed batteries of open-response and multiple-choice questions (referred to as the Kentucky Core Content Tests) in selected contents for each grade.

	EP	4	5	6	7	8	9	10	11	12
Reading		■			■			■		
Mathematics			■			■			■	
Science		■			■				■	
Social Studies			■			■			■	
Arts & Humanities			■			■			■	
Practical Living/ Vocational Studies			■			■		■		
Writing		■			■					■
NRT	■			■			■			

In reading, mathematics, science and social studies, 6 forms of the test were administered, each containing 6 open-response and 24 multiple-choice questions used for reporting and accountability purposes. (Each form also included an additional open-response item and 4 multiple-choice items for field test purposes, bringing the total to 7 open-response and 28 multiple-choice. Field test items are not included in reporting or accountability data.)

In arts & humanities and practical living/vocational studies, there were 12 forms of the assessment, each containing 2 open-response and 8 multiple-choice items used for reporting and accountability purposes. (An additional open-response and 4 multiple-choice items were included for field test purposes.)

Writing data are based on the administration of writing prompts distributed across 6 forms (students select one of two prompts) and the writing portfolio.

Multiple-choice questions are included in the 2002 data reported here and are combined with the open-response data. They are included such that multiple-choice data are weighted at approximately 33% and open-response items at a weight of approximately 67%.

Students in grades end-of-primary, 6 and 9 completed batteries of multiple-choice questions on the CTBS/5 (referred to as the National Norm Referenced Test) in selected content areas of reading, language arts and mathematics.

Schools are held accountable for all of the students enrolled in the school as of the first day of the testing window.

Kentucky law states that, "schools shall expect a high level of achievement of all students." It also states that, "schools shall be rewarded for an increased proportion of successful students, including those students who are at risk of school failure."

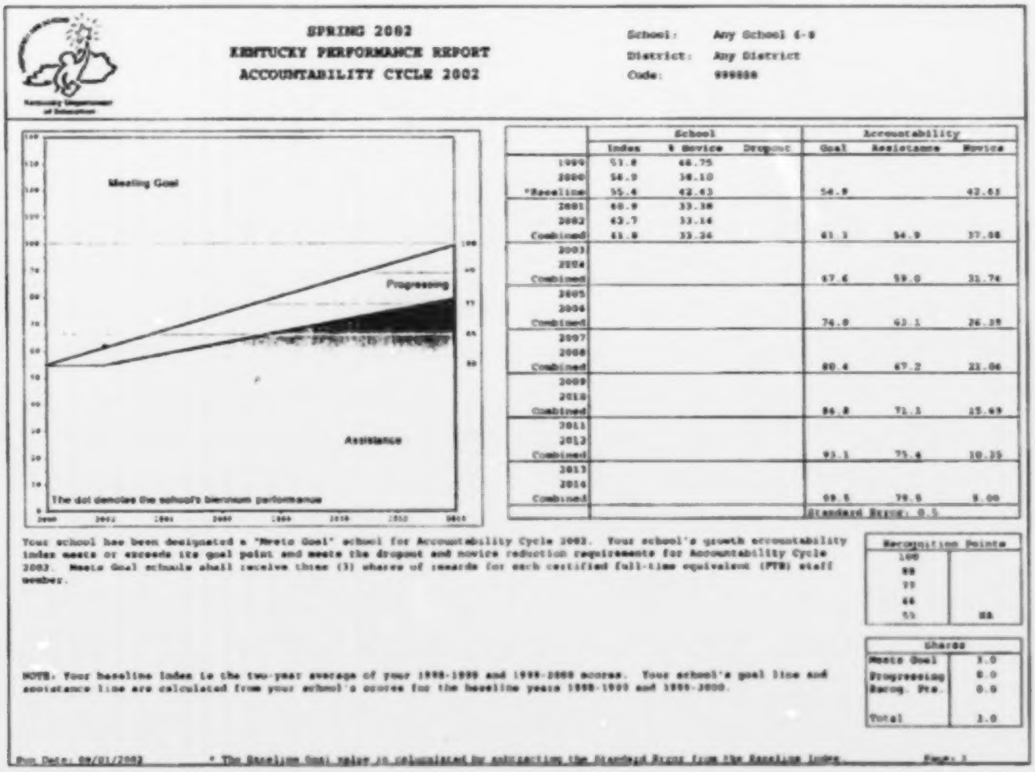
Therefore, there are virtually no exemptions from the testing. Students not included in the data summarized here include:

- Foreign exchange students.
- Students determined to be medically unable to participate in the assessment.
- (at the school's option) Limited English-speaking students who have been enrolled in an English-speaking school for fewer than two years.

The number and percent of students who did not participate for these reasons are provided in this report. Any other student for whom the school is accountable but who was not tested is assigned to the "Novice Non-Performance" level. The number and percentage of students who received this type of "Novice" rating are also in the report.

Accountability Cycle 2002

This page summarizes information pertaining to a school's Accountability Classification. It presents the Growth Chart unique to each school and a table featuring school results and school accountability target values. (See the section above on the Long-Term Accountability Model for more details.) The Growth Chart includes a Goal Line represented by a straight line that begins in 2000 at the Baseline and ends in 2014 at 100. Note that the school in this example has a Baseline of 55.4. The actual "beginning point" for the school is equal to this value minus the standard error, or 55.4 - .5 = 54.9. The Baseline "beginning point" value of 54.9 appears in the table under Accountability, in the first cell of the Goal column. The standard error of 0.5, used to compute the beginning point, also appears under Accountability, in the last cell. Other important target values for the school also appear under Accountability. Each of the three columns under the heading "Accountability" -- Goal, Assistance, and Novice -- represents values used in the Long-Term Accountability System.


Meeting Goal
Progressing
Assistance
The dot denotes the school's biennium performance

SPRING 2002
KENTUCKY PERFORMANCE REPORT
ACCOUNTABILITY CYCLE 2002

School: Any School 6-8
District: Any District
Code: 999888

	School			Accountability		
	Index	% Novice	Dropout	Goal	Assistance	Novice
1999	51.8	46.75				
2000	58.9	38.10				
*Baseline	55.4	42.43		54.9		42.43
2001	60.9	33.38				
2002	62.7	33.14				
Combined	61.8	33.26		61.3	54.9	37.08
2003						
2004						
Combined				67.6	59.0	31.74
2005						
2006						
Combined				74.0	63.1	26.39
2007						
2008						
Combined				80.4	67.2	21.04
2009						
2010						
Combined				86.8	71.3	15.69
2011						
2012						
Combined				93.1	75.4	10.35
2013						
2014						
Combined				99.5	79.5	5.00
				Standard Error: 0.5		

Your school has been designated a "Meets Goal" school for Accountability Cycle 2002. Your school's growth accountability index meets or exceeds its goal point and meets the dropout and novice reduction requirements for Accountability Cycle 2002. Meets Goal schools shall receive three (3) shares of rewards for each certified full-time equivalent (FTE) staff member.

Recognition Points	
100	
88	
77	
66	
55	NA

Shares	
Meets Goal	3.0
Progressing	0.0
Recog. Pts.	0.0
Total	3.0

NOTE: Your baseline Index is the two-year average of your 1998-1999 and 1999-2000 scores. Your school's goal line and assistance line are calculated from your school's scores for the baseline years 1998-1999 and 1999-2000.

Run Date: 09/01/2002 * The Baseline Goal value is calculated by subtracting the Standard Error from the Baseline Index. Page: 1

The other values listed under Goal, 61.3, 67.6, 74.0, 80.4, 86.8, 93.1, and 99.5 are the school's unique targets or goals for each biennium depicted on the Growth Chart. While the Growth Chart represents a useful tool for tracking a school's progress toward Proficiency, it is the values printed under Goal give the precise target a school has to meet or exceed in a biennium to be in the Meeting Goal area of the graphic, and thus on target to reach 100 or Proficiency by 2014.

Also presented in the Assistance column under Accountability are values comprising the Assistance line, i.e., the line separating the Assistance area from the Progressing area on the Growth Chart. In the above example, these values, starting in 2002 are: 54.9, 59.0 63.1, 67.2, 71.3, 75.4, and 79.5. Note that the Assistance point in 2002 (i.e., 54.9) is the same as the 2000 Baseline "beginning point" appearing directly above it. These points were determined by essentially taking the Goal line, sliding it over two years such that the Baseline value was associated with the 2002 biennium, and then tilting that line so that it ended in 80 in 2014 instead of 100. The standard error of 0.5 is subtracted from the Assistance Line, just as it was subtracted from the Baseline; this is why the Assistance Line begins at the Baseline beginning point of 54.9 and ends at 79.5, or 80 - .5 = 79.5. A school falling on or above the Assistance Line, but below the Goal Line, is in the Progressing area, while a school falling below the Assistance Line is in the Assistance area.

The Goal Line and the Assistance Line each incorporate a standard error ranging in size from approximately 0.5 to 3.0 depending upon school level (elementary, middle and high school) and school size. Larger schools with many students will have a smaller standard error than smaller schools with fewer students. On page 20 of this document, the standard error in the Growth Chart is represented by the thickness of the line. In practice, as seen on page 24, the standard error is subtracted first, and then a "thin" line is drawn to depict the Goal Line and the Assistance Line. That is, a fairness margin is included for both lines. The fairness margin takes into account that there are errors of measurement in any assessment program. These errors are not errors in the sense that a mistake has been made; rather, they reflect the realization that measurement is imprecise. Introductory statistics courses teach that measurement error should always be taken into account when interpreting test scores. In fact, measurement experts strongly recommend that test publishers and other reporting agencies properly represent measurement error when reporting test scores. For example, confidence intervals are often built around individual student scores. In providing a standard error or fairness margin for the Goal and Assistance Lines, the Long-Term Accountability Model gives an acceptable cushion to schools in that if a school is just *below* the Goal line, but within one standard error, the school is treated as if (or categorized as if) the school was at or above the Goal Line. The same holds true for the Assistance Line.

Important targets for Novice reduction for each biennium are presented in the third column under Accountability. With regard to Novice reduction, schools must reduce their percent of Novices on a schedule so that by 2014, the school has 5% or less of its students scoring Novice. Under the column labeled "Novice" are the precise Novice reduction targets needed for the school to have only 5% Novice by the year 2014. The Baseline for the Novice reduction criteria was calculated by first obtaining the percent of Novice in each of the seven content areas (i.e., reading, mathematics, science, social studies, arts and humanities, practical living/vocational studies and writing). Each of these percentages was then weighted by the same weights used to calculate an Accountability Index (see page 18). Next, five percent was subtracted from the Baseline percent Novice and the remainder divided by seven (the number of biennia from 2002 to 2014). Finally, this last figure was subtracted from the Baseline value once to determine the Novice reduction goal for 2002, twice to determine the Novice reduction goal for each 2004, three times for 2006 and so on for each of the remaining biennia.

While all the values provided in the three columns under Accountability represent targets established from the baseline years of 1999 and 2000, the data in the three columns under "School" represent *actual* school values for the school years listed in the first column (e.g., 1999, 2000, *Baseline, 2001, 2002, Combined, 2003, 2004, Combined, etc.). For example, the first column labeled "Index" contains the Accountability Indices achieved by the school during the school years listed. In the above example, the school had an Accountability Index of 60.9 for 2001 and 62.7 for 2002. The school's combined Accountability Index for the biennium ending in 2002 was 61.8. It is the value of 61.8 that is compared to the Goal and Assistance points to help determine the school's Accountability Classification. In this example, the combined Accountability Index of 61.8 is greater than or equal to the 2002 goal of 61.3. This places the school above the Goal Line and into the Meeting Goal area of the Growth Chart (see page 24).

The second column under "School" presents the school's percentage of Novices. Note how the school also met this criterion (i.e., 33.26 is less than 37.08). Finally, because the school in this example is a middle school, the Dropout criterion does not apply. Because this school met its target Accountability Index and reduced its percent of Novices, the school is classified as a "Meets Goal" school. The number of Reward Shares (including the shares for passing through Recognition Points) is reported on the bottom, right-hand side of the page. In addition, note that the Accountability Classification for the school is outlined in text messages at the bottom of the report.

In addition to the accountability criteria discussed above, schools can achieve Rewards three other ways as long as the Novice reduction and Dropout criteria have been satisfied:

- If a school is in the Progressing area of the Growth Chart, and increased its Accountability Index in the second biennium, the school is eligible for one-half share of rewards.
- If a school passes any one of the five Recognition Points (i.e., 55, 66, 77, 88, 100) the school is eligible for one share of rewards.
- If a school is in the top five percent of all schools and has met or exceeded the fourth recognition point, the school is eligible for one share of rewards if the school is not receiving rewards any other way.

Besides establishing a system of rewards for school improvement, CATS also provides sanctions for schools that decline by an unacceptable margin (see 703 KAR 5:120 Assistance for schools; guidelines for scholastic audit). According to regulation, all schools falling into the Assistance classification are rank-ordered from highest to lowest according to the school's combined 2001/2002 accountability index. This set of schools is then divided into thirds. The top third are designated Level 1 schools, the middle third Level 2, and the bottom third Level 3. The following bullets briefly summarize the audit/review process for these schools:

- Level 3 Schools will be scheduled for scholastic audits by an external team coordinated by KDE. The school shall adhere to the requirements for a "Level 3" school as defined in 703 KAR 5:120 Sections 4, 5, 6, 7, 8 and 9. Level 3 schools shall receive education assistance from a highly skilled educator under KRS 158.782 and a scholastic audit.

Assistance Level 3 schools may be eligible to receive Commonwealth school improvement funds.

- Level 2 Schools are required to receive a scholastic review by a team set up by KDE. The team must include local district members. The school shall adhere to the requirements for a "Level 2" school as defined in 703 KAR 5:120 Section 3. Level 2 schools shall receive a scholastic review facilitated by a designee of the Commissioner of Education with assistance from the district's central office staff. Assistance Level 2 schools may be eligible to receive Commonwealth school improvement funds.
- Level 1 Schools are required to receive a scholastic self-review by a team set up by the local school district. The school shall adhere to the requirements for a "Level 1" school as defined in 703 KAR 5:120 Section 2. Level 1 schools must conduct a scholastic review and self-study facilitated by the district's professional development coordinator with assistance provided by Kentucky Department of Education staff. Assistance Level 1 schools may be eligible to receive Commonwealth school improvement funds.

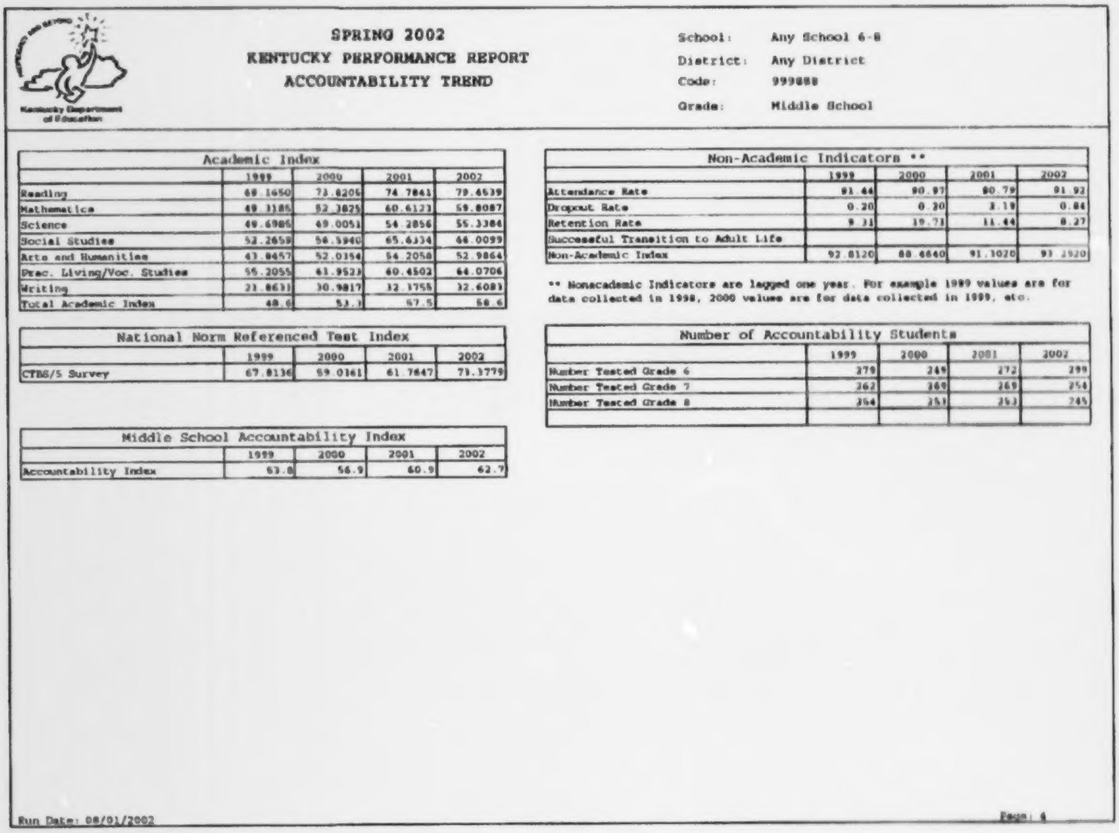
Some important questions school personnel may want to ask pertaining to a school's Accountability Classification include:

- What is the school's accountability goal for 2002?
- Did the school meet its accountability goal?
- Did the school meet its novice reduction and dropout goals?
- What is the baseline for the school? What is the standard error for the school?
- Did the school pass a recognition point?
- Would the school qualify for a reward?
- What is the school's goal for the next biennium?

Accountability Trend

An example of the Accountability Trend page is provided on the following page of this document. The Accountability Trend page provides more detailed summary information relative to a school's accountability calculations for each year of the cycle, including academic indices for each content area, nationally norm-referenced test indices, non-academic indicators and the number of accountability students. While some of the same information on this page is presented in a more graphic, user-friendly format on other pages of the KPR (for example, see Content Area Index Trends section below), the Accountability Trend page is important because it provides a one page look at many aspects of accountability data. For example, this is the only page of the KPR that provides the non-academic data and NRT indices for four years of the accountability cycle.

Not only can the academic index trends across years be evaluated to assess growth to determine the relative strengths and weaknesses in each content area, but also values on this page can be used to replicate or check the calculation of the Accountability Index for each year. For example, the content area index computations include scores of Alternate Portfolio students and are carried out to four decimal places, the same precision used by the Department of Education and its contractors in their calculations. While the Growth Chart on the Accountability Cycle 2002 page of the KPR gives a very global summary of a school's accountability results, the Accountability Trend page provides the next level of detail as one "drills-down" through the data provided in the KPR. The Accountability Trend page allows that first glimpse of what content areas need more attention and which are possible sources for best practice support. In each case, more detailed, content area specific pages of the KPR (discussed below) need to be reviewed. The next page in the sequence of reports addresses the important issue of disaggregation gap trends.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
ACCOUNTABILITY TREND

School: Any School 6-8
District: Any District
Code: 999888
Grade: Middle School

Academic Index	1999	2000	2001	2002
Reading	68.1650	73.8205	74.7861	79.4539
Mathematics	49.3185	52.3825	60.6123	59.8087
Science	49.6986	49.0051	54.2856	55.3384
Social Studies	52.2659	56.5940	65.6334	66.0099
Arts and Humanities	43.8457	52.0354	54.2058	52.9864
Prac. Living/Voc. Studies	55.2055	61.9523	60.4502	64.0706
Writing	23.8631	30.9817	32.3755	32.6083
Total Academic Index	48.6	53.3	57.5	58.6

National Norm Referenced Test Index	1999	2000	2001	2002
CTBS/5 Survey	67.8136	59.0161	61.7647	73.3779

Middle School Accountability Index	1999	2000	2001	2002
Accountability Index	53.8	56.9	60.9	62.7

Non-Academic Indicators **	1999	2000	2001	2002
Attendance Rate	91.44	90.97	90.79	91.92
Dropout Rate	0.20	0.30	3.19	0.84
Retention Rate	9.31	19.71	11.44	8.27
Successful Transition to Adult Life				
Non-Academic Index	92.8120	88.4640	91.3020	93.2920

** Nonacademic Indicators are lagged one year. For example 1999 values are for data collected in 1998, 2000 values are for data collected in 1999, etc.

Number of Accountability Students	1999	2000	2001	2002
Number Tested Grade 6	279	249	272	299
Number Tested Grade 7	262	269	269	254
Number Tested Grade 8	264	253	253	245

Run Date: 08/01/2002

Page: 4

Some important questions school personnel may want to ask include:

- Did any academic areas show steady growth over the years? If yes, which areas and more importantly, why?

- Did any academic areas decline or show inconsistent performance over the years? If yes, which areas and why?
- Did any of the non-academic data show movement in either a positive or negative direction? Explain.
- Does the NRT data show change? If so, how.

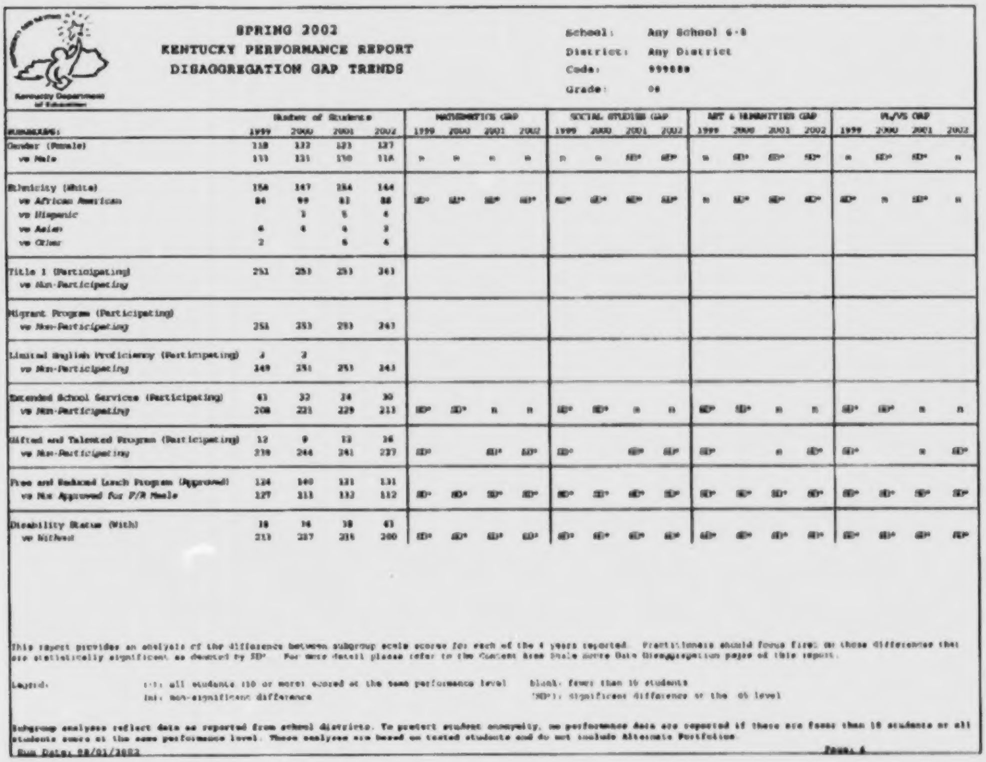
Disaggregation Gap Trends

An example of the Disaggregation Gap Trends page is provided on the following page of this document. Depending upon school configuration, this report contains one to two pages that summarize scale score differences between certain student groups across multiple years of the assessment. The report is new this year and has been included in the KPR to provide the equity analysis required under SB 168 (closing achievement gaps). A test of statistical significance is given for each comparison for each year. The number of students contributing to the calculation of each significance test is also reported[1]. The scale scores and the numerical value of the gap between scale scores for student groups are *not* reported on this page of the KPR but are reported for 2002 on the Scale Score Data Disaggregation pages (see Scale Score Data Disaggregation section below). These same values for earlier years of the KPR can be obtained from the KDE website. The gap is assessed for each content area by gender, ethnicity, Title 1 programs, migrant programs, students with limited English proficiency, Extended School Service programs, gifted and talented programs, students participating in free or reduced price lunch versus those students not participating in free or reduced price lunch, vocational education (high school only), and students with and without disabilities.

The Disaggregation Gap Trends page provides a very global summary of the comparison between important student groups. Gap values that are statistically different beyond the .05 level of significance are "flagged" by the notation "SD*". Differences that were found to be statistically non-significant are denoted by an "n". A statistically significant difference between any two student groups represents the starting point for further investigation of the difference. It is important for schools to follow-up on any significant differences. Strategies available for this follow-up include (a) assessing the scale scores to determine the magnitude of the difference as well as how much each of the groups must gain to reach the next performance level cut point, (b) using the data disaggregation provided on the KPR to further study the percentage of Novice, Apprentice, Proficient, and Distinguished students for each student group, and (c) consideration of the number of students making up each group and the fact that some students may be in more than one group (e.g., males who have a disability).

---

[1] For more information on scale scores and how the significance test statistics were calculated, see the Scale Score Data Disaggregation section below.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
DISAGGREGATION GAP TRENDS

School: Any School 6-8
District: Any District
Code: [illegible]
Grade: [illegible]

Some important questions school personnel may want to ask include:

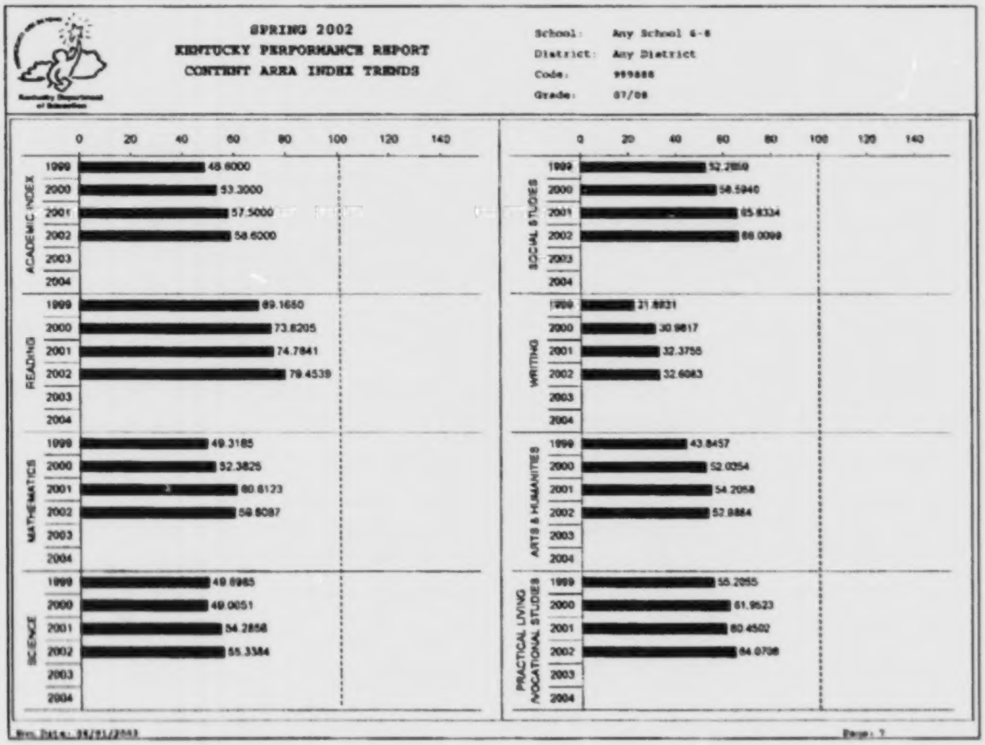
- Where are the significant differences for the school (hint: note the SD*'s and n's)?
- What groups in which content areas show significant differences for three or more years?
- Are there groups in content areas that show no significant differences for any years?
- Where in the KPR can additional details about disaggregation be found?
- What is the importance of the n counts (i.e., the number of students)?

Two cautionary notes should be kept in mind when reviewing disaggregated data for schools: (a) the accuracy of the disaggregated data is dependent on how schools filled in this information on the Student Response Booklets and (b) if fewer than ten students were reported in a school or district for a category, or there were more than ten students but all students scored at the same performance level (i.e., Novice, Apprentice, Proficient, or Distinguished), no analysis of the gap was provided to ensure the protection of the privacy of individual students. With these cautions in mind, data disaggregation information can be helpful to schools and districts in evaluating student performance in relation to special educational programs, e.g., Title 1, Extended School Services (ESS). This information can also be used in consolidated planning to address issues relevant to equity across diverse student groups.

The Title 1 disaggregation has a few characteristics unique to the Title 1 program, which need to be noted. If a school participates in a school-wide Title 1 program, the disaggregation of student performance is for all students in the school. If a school participates in a Title 1 Targeted Assistance program, only the students participating in this program are part of the disaggregation data (as indicated by school staff on the Student Response Booklet).

Content Area Index Trends

An example of the Content Area Index Trends page is provided below. This one page report gives comparisons/trends across multiple years *within* each content area and for the overall academic index. Horizontal bar charts are used to compare data across the years and a separate page is provided for each level (i.e., elementary, middle and high school) where necessary. Indices are graphed beginning with the spring 1999 Kentucky Core Content Test. Index values are printed next to each bar and reflect the 0 to 140 scale. It should be noted that each index value includes the scores of students participating in the alternate portfolio. Values for each year and content area are rounded to four decimal places and can be used to replicate the calculation of accountability indices for each year. Please note that comparisons should only be made within a content area and *not* across content areas. Identical index values across content areas may have different instructional implications.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
CONTENT AREA INDEX TRENDS

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07/08

Some important questions school personnel may want to ask include:

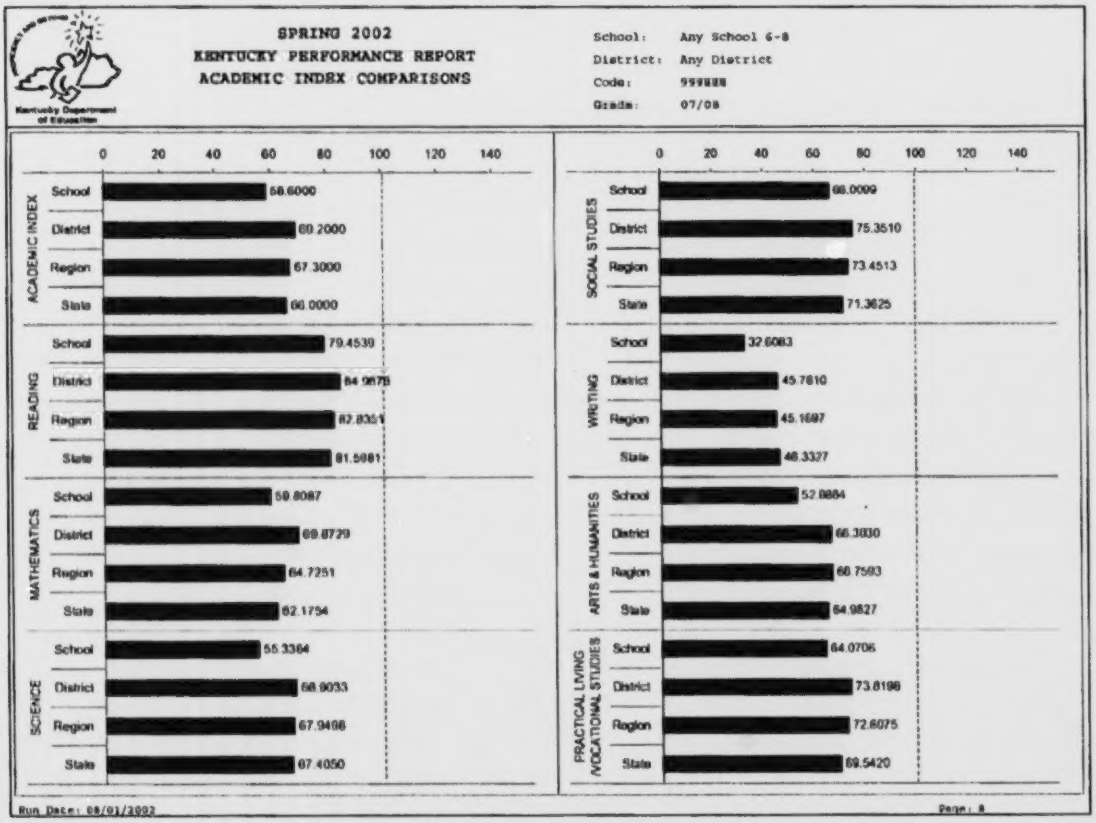
- Are there any content areas that declined over the years, were flat, or showed uneven performance?
- How does each content area compare to the absolute standard of 100? How close is the academic index to 100? How close is each content area?
- Did any content area show consistent growth?
- What questions could be asked of teachers and others in a school to identify possible causes for the patterns that appear in the scores?
- What instructional targets might someone suggest?

Academic Index Comparisons

An example of the Academic Index Comparisons page is presented on the following page. The Academic Index Comparisons report provides a one-page comparison of school, district, region and state academic indices for each content area and for the overall academic index used in accountability. A separate page is provided for each grade level (i.e., elementary, middle and high school). For each index, comparisons are made using horizontal bars stacked one below the other in the following order: school, district, region and state. Index values are printed next to each bar and reflect the 0 to 140 scale. For the academic index and each content area index, the four bars provide a visual comparison of the *current* year standing of the school as compared to the school's district, region and the state. As such, the comparisons provided on this page (e.g., the difference between the school and region) should be interpreted as normative.

While comparisons among levels are normative, index values for the school are the same values used for calculating the school's academic, and thus, accountability index. Because of this, the school indices provide an indication of how close a school is to the state goal of 100 (i.e., Proficiency) by 2014. The district, region and state indices also provide an indication of how close each is to the state goal of 100. Note that specific content area index values are reported to four decimal places so academic index calculations can be verified/replicated. The overall academic index values are reported to one decimal place.

The comparisons provided on this page of the KPR can give a preliminary indication of which academic content areas are strong and which may require additional attention. Although the state goal for all schools is to have a combined 2013/2014 accountability index of 100, school content area indices that are considerably lower than the region or state, especially relative to other content area indices, should be further studied to determine possible reasons/solutions for why the indices are lower. In other words, the Academic Index Comparisons page provides a global, first look at your school's indices. Other report pages on the KPR will have to be referenced to gain more detailed information on the performance of students in your school. The content area Trends Data, Number and Percent page (see below) can provide this additional information.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
ACADEMIC INDEX COMPARISONS

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07/08

Run Date: 08/01/2002

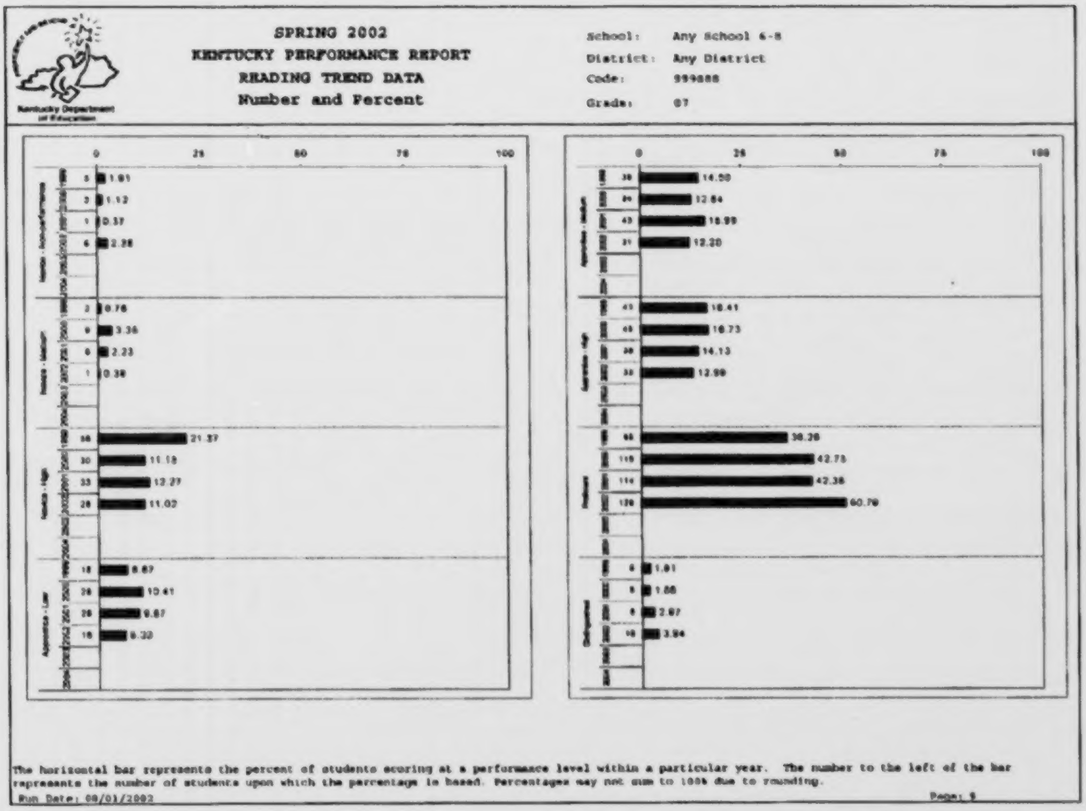
Page: 8

Some important questions school personnel may want to ask include:

- In what content area(s) is the school out performing the district, region and state?
- In what content area(s) is the school performing lower than the district, region and state?
- What perspective can this give to the school? Note: Remember to compare to the absolute goal of 100.

Trend Data, Number and Percent

This page begins the "cluster" of reports for each content area. For a content area (e.g., reading), a single page gives horizontal bar charts for year-to-year comparisons of the percentage of students achieving Distinguished, Proficient, Apprentice (high, medium and low) and Novice (high, medium and non-performance). This data can be used to help schools assess their strengths and weaknesses in each content area including how their students are progressing through the Novice and Apprentice performance levels.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
READING TREND DATA
Number and Percent

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07

The horizontal bar represents the percent of students scoring at a performance level within a particular year. The number to the left of the bar represents the number of students upon which the percentage is based. Percentages may not sum to 100% due to rounding.

Run Date: 08/01/2002 Page: 5

An example of this report is presented above. One page of trend data is provided for each content area (reading, mathematics, science, social studies, arts and humanities and practical living/vocational studies) and includes comparisons across four years (1999, 2000, 2001 and 2002). The horizontal bar charts give a visual comparison of percentages across multiple years of the assessment. Note that the percentages are printed at the end of each bar and are given to two decimal places so that content area academic index calculations can be verified/replicated.

Because the goal of 100 (Proficiency) for schools can be reached by reducing the number/percent of students scoring Novice, and increasing the number/percentage of students scoring Proficient and Distinguished, the Novice bars (high, medium and non-performance) should steadily decrease in size as one views the chart across years, while the Proficient and Distinguished bars should steadily increase in size across the years. Weather or not these two separate goals are being achieved by a school is readily seen by simply viewing the bars across subsequent years.

The trend data for writing has two pages because writing performance is evaluated two ways in CATS: the Writing Portfolio and the on-demand writing prompt. Each of these pages displays the same information on performance levels as in the other content areas. Note that for Writing Portfolios, scoring was done by teachers who scored the portfolios at the school level or by audit scorers if the school participated in the Writing Portfolio audit at grades 4, 7 and 12. The assessment contractors scored the on-demand writing prompt.

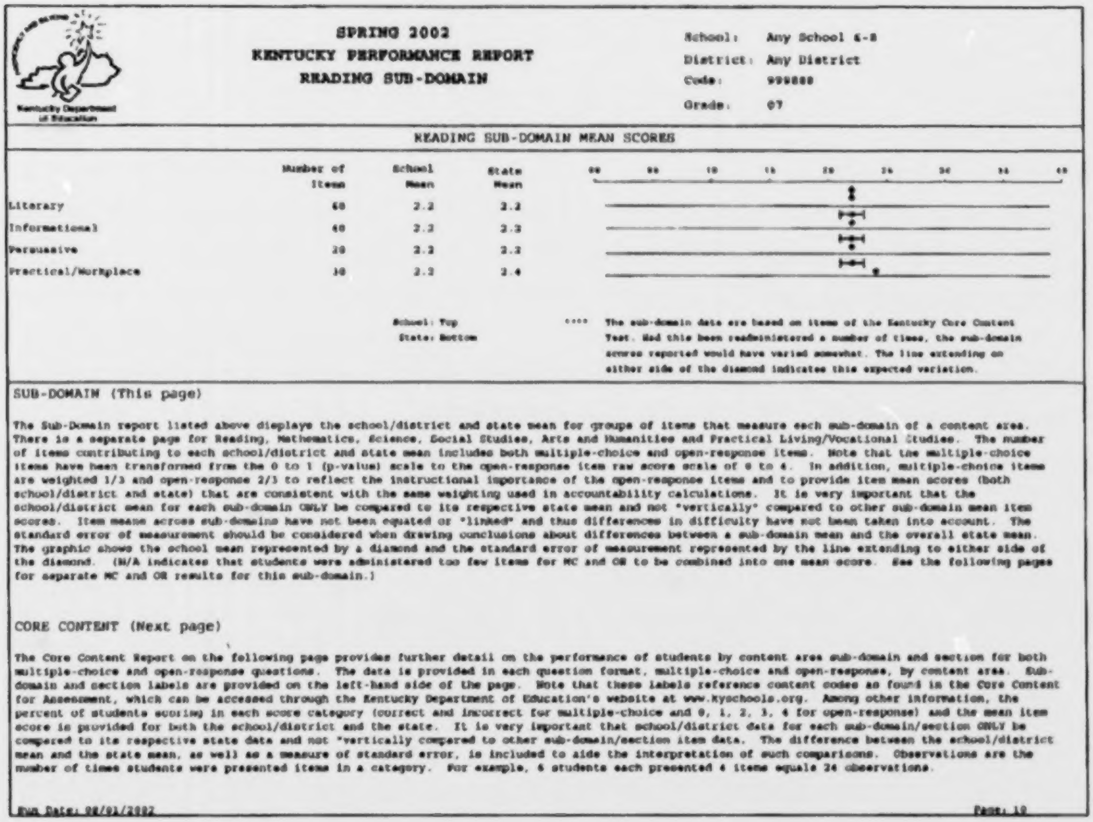
Some important questions school personnel may want to ask include:

- How has the percent of students in each category changed over time?
- Has the percent increased in the upper levels (Proficient/Distinguished) and decreased in the lower levels (Novice/Apprentice)?
- What does the data show about students in the lowest performance levels?
- How might this information impact the Novice reduction rule reflected on the Growth Chart?
- What kind of emphasis or programs would the school need to consider implementing to change the current pattern of results?

Sub-Domain

An example of the Sub-Domain report, the second page of the "cluster" of reports for each content area, is provided on the following page. The Sub-Domain report displays the school and state mean for groups of items that measure each sub-domain of a content area. There is a separate page for reading, mathematics, science, social studies, arts and humanities and practical living/vocational studies. The number of items contributing to each school and state mean includes both multiple-choice and open-response items. Note that the multiple-choice items have been transformed from the 0 to 1 (p-value) scale to the open-response item raw-score scale of 0 to 4. In addition, multiple-choice items are weighted 1/3 and open-response 2/3 to reflected the instructional importance of the open-response items and to provide item mean scores (both school and state) that reflect the same weighting used in accountability calculations. *It is very important that the school mean for each sub-domain* ***ONLY*** *be compared to its respective state mean and not "vertically" compared to other sub-domain mean item scores.* Item means across sub-domains have *not* been equated or "linked" and thus differences in difficulty have not been taken into account. The standard error of measurement, denoted by the bar running through the school mean, should be considered when drawing conclusions about differences between a sub-domain mean and the overall state mean.

The mean item scores can be used to identify sub-domain areas a school may want to target for future improvement. In the example below, the school mean and state mean are the same for each sub-domain except for the last, practical/workplace, where the school mean of 2.2 is lower than the state mean of 2.4. Because the school standard error "bar" does not overlap the state mean, the difference between the two values can be considered important enough to warrant further examination. The Core Content pages of the KPR (formally a separate report produced by KDE) discussed in the next section can provide that further insight into the strengths and weaknesses of a content area.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
READING SUB-DOMAIN

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07

READING SUB-DOMAIN MEAN SCORES

	Number of Items	School Mean	State Mean
Literary	60	2.2	2.2
Informational	40	2.3	2.3
Persuasive	20	2.3	2.3
Practical/Workplace	10	2.3	2.4

School: Top
State: Bottom

The sub-domain data are based on items of the Kentucky Core Content Test. Had this been readministered a number of times, the sub-domain scores reported would have varied somewhat. The line extending on either side of the diamond indicates this expected variation.

SUB-DOMAIN (This page)

The Sub-Domain report listed above displays the school/district and state mean for groups of items that measure each sub-domain of a content area. There is a separate page for Reading, Mathematics, Science, Social Studies, Arts and Humanities and Practical Living/Vocational Studies. The number of items contributing to each school/district and state mean includes both multiple-choice and open-response items. Note that the multiple-choice items have been transformed from the 0 to 1 (p-value) scale to the open-response item raw score scale of 0 to 4. In addition, multiple-choice items are weighted 1/3 and open-response 2/3 to reflect the instructional importance of the open-response items and to provide item mean scores (both school/district and state) that are consistent with the same weighting used in accountability calculations. It is very important that the school/district mean for each sub-domain ONLY be compared to its respective state mean and not "vertically" compared to other sub-domain mean item scores. Item means across sub-domains have not been equated or "linked" and thus differences in difficulty have not been taken into account. The standard error of measurement should be considered when drawing conclusions about differences between a sub-domain mean and the overall state mean. The graphic shows the school mean represented by a diamond and the standard error of measurement represented by the line extending to either side of the diamond. (N/A indicates that students were administered too few items for MC and OR to be combined into one mean score. See the following pages for separate MC and OR results for this sub-domain.)

CORE CONTENT (Next page)

The Core Content Report on the following page provides further detail on the performance of students by content area sub-domain and section for both multiple-choice and open-response questions. The data is provided in each question format, multiple-choice and open-response, by content area. Sub-domain and section labels are provided on the left-hand side of the page. Note that these labels reference content codes as found in the Core Content for Assessment, which can be accessed through the Kentucky Department of Education's website at www.kyschools.org. Among other information, the percent of students scoring in each score category (correct and incorrect for multiple-choice and 0, 1, 2, 3, 4 for open-response) and the mean item score is provided for both the school/district and the state. It is very important that school/district data for each sub-domain/section ONLY be compared to its respective state data and not "vertically compared to other sub-domain/section item data. The difference between the school/district mean and the state mean, as well as a measure of standard error, is included to aide the interpretation of such comparisons. Observations are the number of times students were presented items in a category. For example, 6 students each presented 4 items equals 24 observations.

Run Date: 08/01/2002 Page: 10

Some important questions school personnel may want to ask include:

- In what sub-domain is the school above or below the state mean?
- What implications exist for instruction and curriculum alignment?
- Why must the information in this report be read horizontally?

Core Content

The Core Content pages of the KPR, the third page of the "cluster" of reports for each content area, replaces the separate Kentucky Core Content Report (KCC) received by schools and districts last year. An example of the reading report is presented on page 38. The format of the report is unchanged from last year and provides further detail on the performance of students by content area sub-domain and section for both multiple-choice and open-response questions. While the data provided for each question format (i.e., multiple choice and open response) is very similar, the data for each is presented on separate pages. Sub-domain and section labels are provided on the left-hand side of the page. Note that these labels reference content codes as found in the *Core Content for Assessment*. Among other information, the percent of students scoring in each score category (correct and incorrect for multiple choice and 0, 1, 2, 3, 4 for open response) and the mean item score across items within the category is provided for both the school/district and the state. *Note that all school/district comparisons within a sub-domain or section must be made with respect to the state's performance within the same content area sub-domain or section.* The difference between the school mean and the state mean, as well as a measure of standard error, is included to aide interpretation of the comparisons.

The *Core Content for Assessment* is organized in the following manner:

- Content area (e.g., MATHEMATICS)
- Sub-domain (e.g., 1.x.x - Number/Computation)
- Section (e.g., 1.1.x – Concepts; 1.2.x – Skills; 1.3.x – Relationships)
- Bullet (not provided on KPR at this level)

For example, for mathematics, the Core Content codes are:

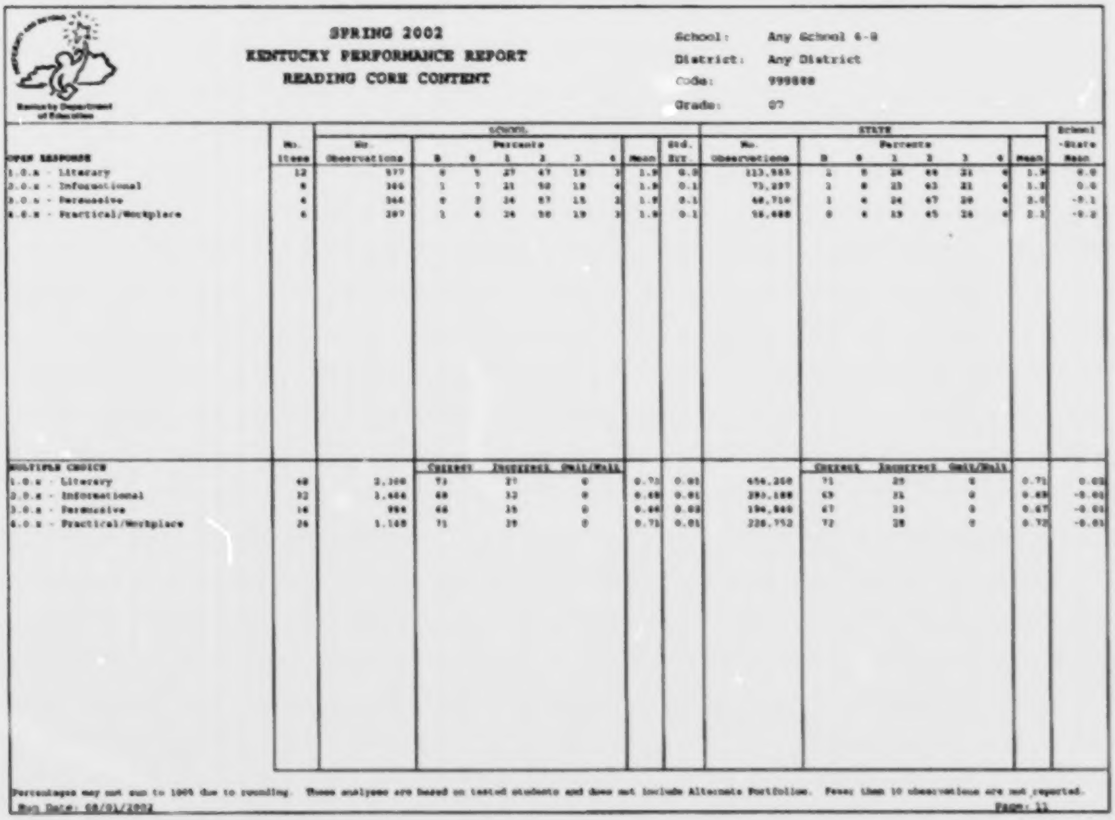
MATHEMATICS:

1.x.x - Number/Computation
1.1.x - Concepts
1.2.x - Skills
1.3.x – Relationships

2.x.x - Geometry/Measurement
2.1.x - Concepts
2.2.x - Skills
2.3.x – Relationships

3.x.x - Probability/Statistics
3.1.x - Concepts
3.2.x - Skills
3.3.x – Relationships

4.x.x - Algebraic Ideas
4.1.x - Concepts
4.2.x - Skills
4.3.x – Relationships



SPRING 2002
KENTUCKY PERFORMANCE REPORT
READING CORE CONTENT

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07

During test development, Kentucky teachers come together in Content Advisory Committees (CACs) to both write and eventually select items for the Kentucky Core Content Tests. These committees generally include around eight to ten teachers per content area per assessed grade level. The content codes in the *Core Content for Assessment* are applied to specific items during the development process. In other words, Kentucky teachers literally must come to an agreement with respect to the specific content an item measures on the KCCT. As such, this report shows how students performed on specific areas linked directly to the Core Content. Informal feedback to the Department suggests that principals and teachers find this report very useful for evaluating content alignment and instructional practices.

The main features of the report include:

- The number of items that measured the specific area.
- The number of times students were presented items (or had an opportunity to respond to items) in a category (labeled "No. Observations" for number of observations). For example, six students each presented four items equals 24 observations.

- The percent of students scoring in each score category (correct and incorrect for multiple choice and B, 0, 1, 2, 3, 4 for open response).
- The mean item score across items within the specific area for both the school/district and the state. The mean score ranges from 0.00 to 1.00 for multiple choice and from 0.0 to 4.0 for open response.
- In the State section, the difference between the school mean and the state mean is calculated.

Some important strategies school personnel may want to consider include[2]:

- Look at your School Mean column to see which area is the lowest. You might ask questions like: What is the definition of this topic? For example, objects in the sky – how is this defined? Is there a reason this should be the lowest area? Is this an area we teach? How do we teach these topics? What is expected of students in the classroom?
- Compare the school mean with the state mean and look for differences (look at the last column: School - State Mean). Where do you have negative values that are greater than the standard error? What is significant? You could think of a .4 or higher as significant; however, the difference might be relative to each school. For instance, if my school was a full point above the state in all areas except one, I would probably concentrate on that one area even if it was only different by .1.
- Review your percentages of B and 0. Compare these percentages to the state percentages. A score under B indicates a blank answer while a score under 0 indicates answers that were pretty far off task. Are there items that really show up with large percentages of B or 0s? If yes, what is the definition of the item? Is there a reason this item should be this difficult? How do we teach these topics? What is expected of the student in the classroom? How do we assess content like this?
- Look for school means that are high. These areas are places where students did very well. What is the definition of these items? Is there a reason why students did so well? How do we teach these topics? What is expected of the student in the classroom? How do we assess? What implications does this report have for curriculum alignment?

Several cautions to consider while using the Core Content pages of the KPR include:

- Always check the number of test items that measure a Core Content area. Two things may be happening. First, some items are counted as measuring more than one Core Content area. Second, items may be coming from just one or two forms of the test. It's best to remember that some scores come from a limited number of items and a limited number of students.

---

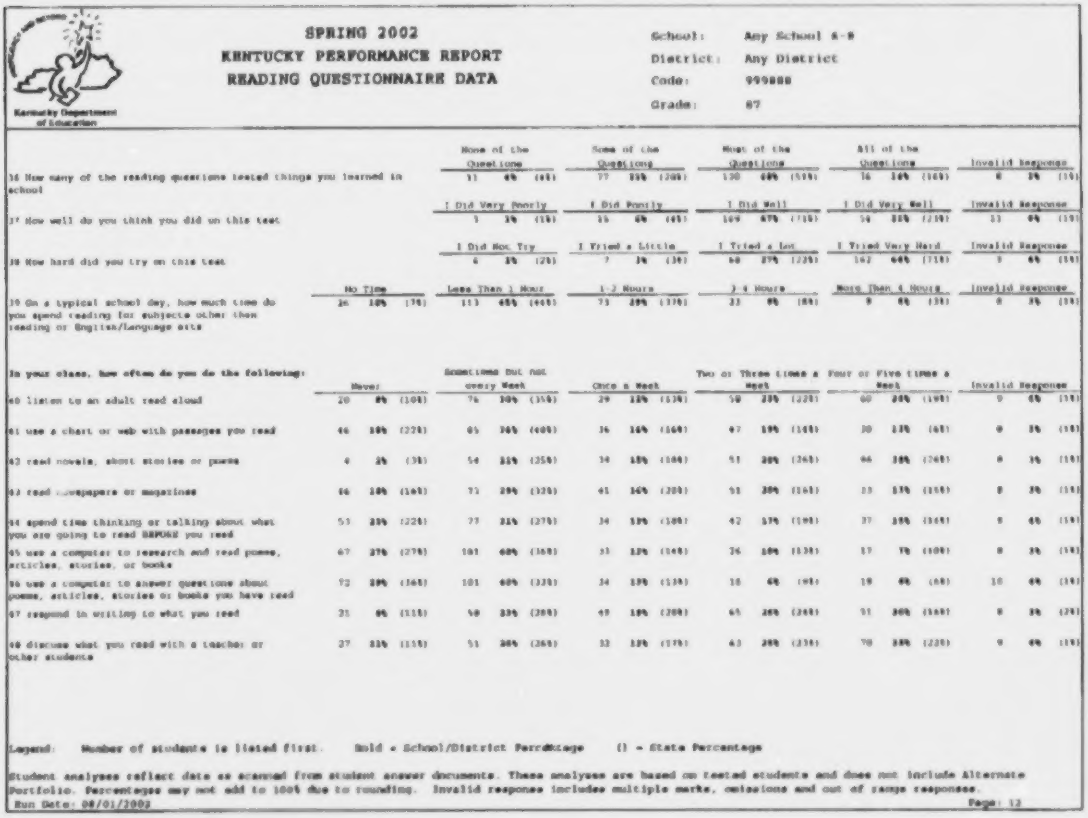
[2] The Department would like to thank Ken Draut for his contribution to the Core Content section of this Interpretive Guide.

- Teachers have a full year perspective on student's ability and the content taught. Teachers' professional judgment should always be taken into account when analyzing test scores.

- Before making any final decisions about curriculum and instruction, please take into account multiple sources of data and ideas. It would be unwise to make any decisions based on one piece of data. Use this report in conjunction with other insights and data.

Questionnaire Data

In addition to the academic questions, students answered a number of questionnaire items. The fourth page of the "cluster" of reports for each content area provides student questionnaire data relevant to the content area. All questionnaire information is based on students who actually answered the questionnaire and may not represent all students who took the test. Questionnaire responses can be useful for studying students' perspective about instructional practices.



Kentucky Department of Education

SPRING 2002
KENTUCKY PERFORMANCE REPORT
READING QUESTIONNAIRE DATA

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07

36 How many of the reading questions tested things you learned in school

37 How well do you think you did on this test

38 How hard did you try on this test

39 On a typical school day, how much time do you spend reading for subjects other than reading or English/Language arts

In your class, how often do you do the following:

40 listen to an adult read aloud

41 use a chart or web with passages you read

42 read novels, short stories or poems

43 read newspapers or magazines

44 spend time thinking or talking about what you are going to read BEFORE you read

45 use a computer to research and read poems, articles, stories, or books

46 use a computer to answer questions about poems, articles, stories or books you have read

47 respond in writing to what you read

48 discuss what you read with a teacher or other students

Legend: Number of students is listed first. Bold = School/District Percentage () = State Percentage

Student analyses reflect data as scanned from student answer documents. These analyses are based on tested students and does not include Alternate Portfolio. Percentages may not add to 100% due to rounding. Invalid response includes multiple marks, omissions and out of range responses.

Run Date: 08/01/2002 Page: 12

Use the Legend at the bottom of the page for aide in understanding the report. Basically, three values are given for each response category. The first value is the number of students who responded to a question in a particular category (e.g., Sometimes but not every week, Once a week, etc.). The second value is **bolded** and gives the percent response for the school. The third value given in parenthesis ( ) is the same percent but for the state. Note that responses under the "Invalid Response" column are for students who did not mark an answer, marked an out-of-range response or marked more than one answer to a question.

Some important questions school personnel may want to ask include:

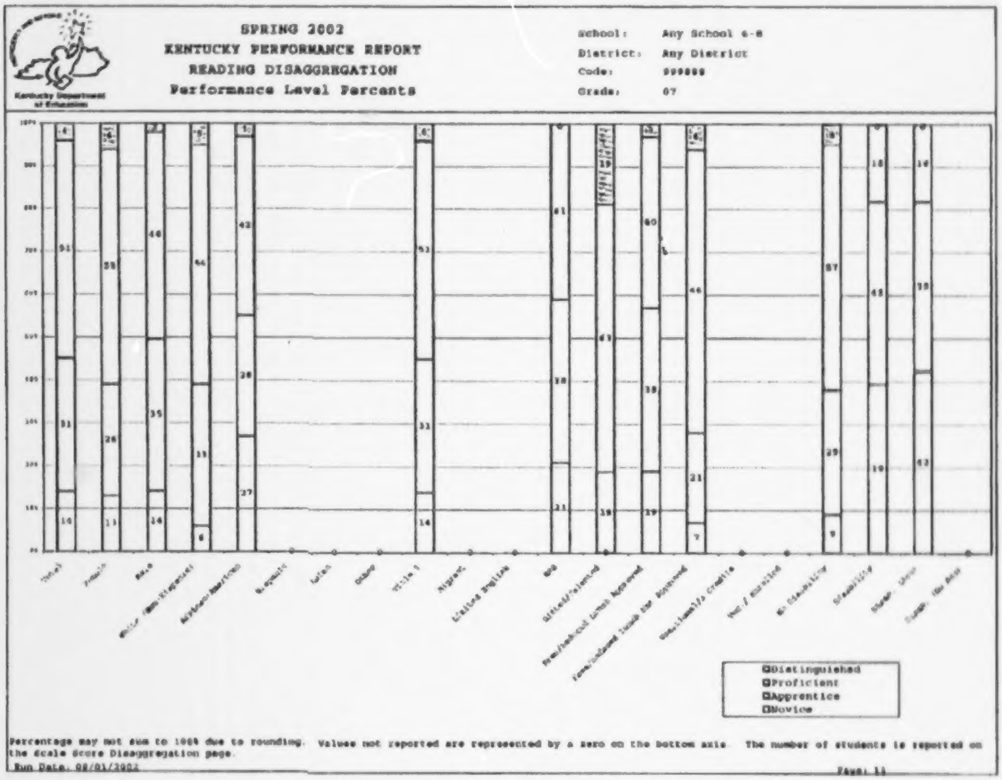
- Are there any notable differences between the school and state percentages?
- Are there implications for different teacher strategies or instruction?
- What questions would you ask in the school to probe deeper?
- What might be some next steps if students and teachers do not share the same perception of instruction?

Disaggregation, Performance Level Percents

The fifth page of the "cluster" of reports for each content area provides stacked bar charts presenting a side-by-side comparison of the percentage of students achieving Distinguished, Proficient, Apprentice and Novice for a number of important student groups. KCCT data are disaggregated based upon gender, ethnicity, Title 1 programs, migrant programs, students with limited English proficiency, Extended School Service programs, gifted and talented programs, students participating in free or reduced price lunch, students not participating in free or reduced price lunch, vocational education (high school only), students with disabilities, students with disabilities receiving accommodations/no accommodations and student without disabilities. These pages of the KPR provide schools and districts with a Data Disaggregation of student performance based on the demographic data requested about each student in their Student Response Booklet.

An example of the Disaggregation, Performance Level Percents page is presented on the following page. One page of stacked bar charts is provided for each content area (Reading, Mathematics, Science, Social Studies, Arts and Humanities and Practical Living/Vocational Studies). The stacked bar charts present a side-by-side comparison of the percentage of students achieving Distinguished, Proficient, Apprentice and Novice for the student groups previously noted. The graphs produced for each content area provide a powerful representation of how each student group is performing on the assessment compared to other student groups. If large differences exist, especially with respect to the percentage of Novice students, the differences are clearly visible upon inspection of the graphs. As such, this series of stacked bar charts may be useful for communicating disaggregation data not only to school personnel, but also to other stakeholder groups, including parents and business leaders.

Two cautionary notes should be kept in mind when reviewing disaggregation data for schools: 1) the accuracy of the disaggregated data is dependent on how schools filled in this information on the Student Response Booklets and 2) if fewer than ten students were reported in a school or district for a category, or more than ten students scored in a category but all these students scored at the same performance level (e.g., all were Apprentice), no disaggregated data was provided to ensure the protection of the privacy of individual students. With these cautions in mind, data disaggregation information can be helpful to schools and districts in evaluating student performance in relation to special educational programs, e.g., Title 1, Extended School Services (ESS). This information can also be used in consolidated planning to address issues relevant to equity across diverse student groups.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
READING DISAGGREGATION
Performance Level Percents

School: Any School 6-8
District: Any District
Code: 999999
Grade: 07

Percentage may not sum to 100% due to rounding. Values not reported are represented by a zero on the bottom axis. The number of students is reported on the Scale Score Disaggregation page.

Run Date. 08/01/2002

The Title 1 disaggregation has a few characteristics unique to the Title 1 program, which need to be noted. If a school participates in a school-wide Title 1 program, the disaggregation of student performance is for all students in the school. If a school participates in a Title 1 Targeted Assistance program, only the students participating in this program are part of the disaggregation data. The district report disaggregates data for all students who participate in either a school-wide or targeted assistance Title 1 program in any school in the district.

Some important questions school personnel may want to ask include:

- Are there any subgroups that have a different pattern? Discuss any pattern(s).
- Are there content areas where a specific student group is showing lower performance than in other areas?
- What implications for opportunity to learn could be discussed from this report page?

Mean Scale Score/Standard Deviation

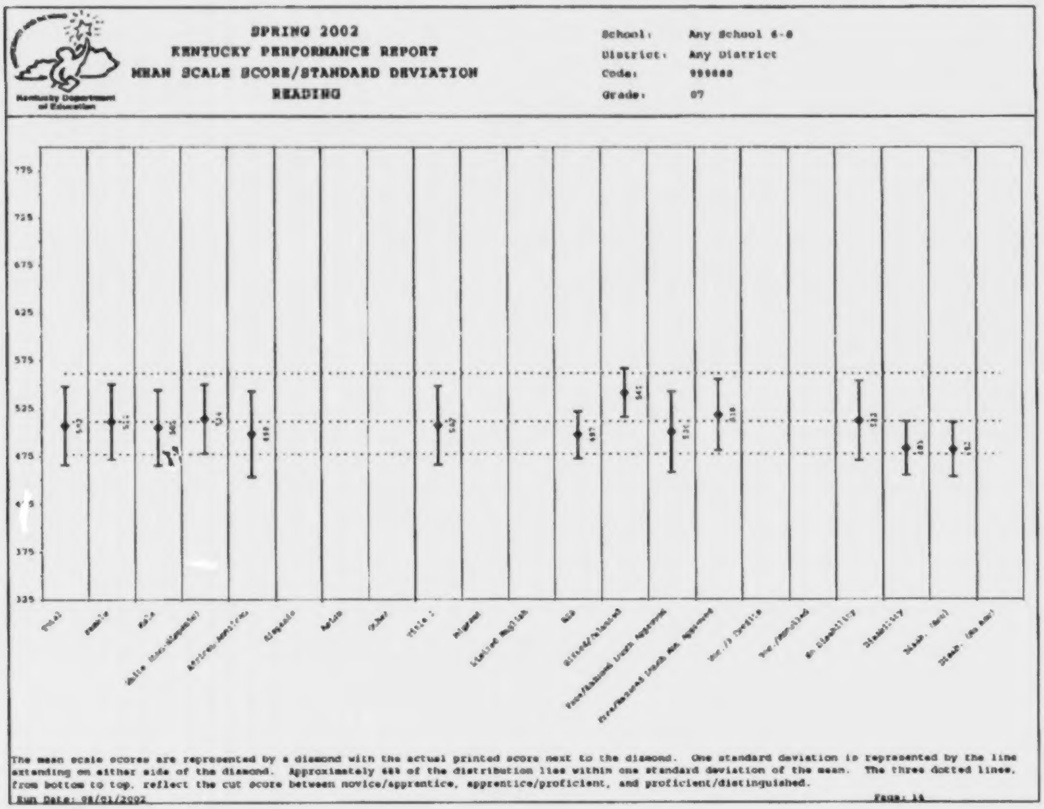
The sixth page of the "cluster" of reports for each content area provides descriptive statistics for scale scores. Scale score means and standard deviations (presented graphically as an interval) are given for a number of important student groups. An example of the Mean Scale Score/Standard Deviation page is presented on the next page of this Interpretive Guide. One page of descriptive statistics is provided for each content area (Reading, Mathematics, Science, Social Studies, Arts and Humanities and Practical Living/Vocational Studies).

Basic descriptive statistics usually involve a measure of central tendency (e.g., mean, median or mode) and a measure of dispersion (e.g., standard deviation or variance). The scale score arithmetic mean and standard deviation are given for the same student groups reported on other pages of the KPR. More specifically, a dot representing the scale score mean (vertical axis) is plotted for each student group (e.g., females, males). Surrounding each dot or scale score mean is an interval that represents one standard deviation below the mean and one standard deviation above the mean, or approximately 68% of students in the group. This representation of scale score means and standard deviations provides a visual summary of the distribution of scores for each student group, side-by-side. If useful, one can actually visualize, or superimpose, a bell shaped curve over each graphed dot and interval, thus taking notice that the graphed values do represent student distributions of scale scores.

On the vertical axis, each of the horizontal lines going across the page is located at a scale score point that represents a performance standard cut point. Recall that this can be done because one page of descriptive statistics is provided for each content area and grade assessed. For example, one "reference" line is drawn across the page for the Novice/Apprentice cut point, one line for the Apprentice/Proficient cut point, and one line for the Proficient/Distinguished cut point. Note that separate lines could also be drawn for the Novice and Apprentice cut points that provide incremental credit (e.g., Apprentice Low, Medium and High). One possible activity for teachers is to draw these additional lines to become more familiar with cut-points and where their students scored in relation to the performance categories. Viewing these "reference" lines across the page provides a strong visual for where the distribution of scores falls for each student group in relation to the state's student performance standards, and can provide direction for where resources need to be focused. The KCCT cut points for each content area and grade level, as well as the Descriptions of performance standards for each content area and grade level, can be found on the Kentucky Department of Education's website at http://www.kde.state.ky.us/.

Some important questions school personnel may want to ask include:

- Which student group(s) has a mean that is close to a cut-point line?
- What implications does this report have for curriculum and instruction?
- How could a school begin to prioritize instructional services to address student needs?


SPRING 2002
KENTUCKY PERFORMANCE REPORT
MEAN SCALE SCORE/STANDARD DEVIATION
READING

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07

The mean scale scores are represented by a diamond with the actual printed score next to the diamond. One standard deviation is represented by the line extending on either side of the diamond. Approximately 68% of the distribution lies within one standard deviation of the mean. The three dotted lines, from bottom to top, reflect the cut score between novice/apprentice, apprentice/proficient, and proficient/distinguished.

Run Date: 08/01/2002 Page: 14

Scale Score Data Disaggregation

On the seventh and last page of the "cluster" of reports for each content area, scale score comparisons are provided for a number of important student groups. A standard error accompanies each scale score. In addition, differences are calculated between the scale scores for certain student groups (e.g., male vs. female, White vs. African-American) and a test of statistical significance is provided for each comparison. Examples of the Scale Score Data Disaggregation pages are presented on page 46 of this Interpretive Guide. These pages of the KPR provide important comparisons between the scale scores of the same student groups reported elsewhere in the KPR.

Mean scale scores for each assessed content area are provided by gender, ethnicity, Title 1 programs, migrant programs, students with limited English proficiency, Extended School Service programs, gifted and talented programs, students participating in free or reduced price lunch, students not participating in free or reduced price lunch, vocational education (high school only), students with disabilities, students with disabilities receiving accommodations/no accommodations and students without disabilities. These scale score means are on the same 325 to 800 scale used for establishing performance standards; the same scale score scale on which cut-points are used for determining Novice, Apprentice, Proficient and Distinguished. As such, the scale score means provided for each content area could be used for analyzing how close (or far) a particular student group is from the next highest performance level.

Accompanying each scale score mean on these data disaggregation pages is a measure of standard error. Standard error values are given in parentheses ( ) next to each mean scale score. These standard error values represent the standard error of the mean for the school and are calculated as:

$$\mathrm{SE} = \frac{sd}{\sqrt{N}} \qquad (1)$$

Where:

SE is the standard error of the school mean,
sd is the standard deviation associated with the scale score mean, and
N is the number of students who took the content area test for a particular grade.

The standard errors (SE) presented on this report are important because they remind us that measurement error should be taken into account when interpreting test scores. For example, if the scale score mean for males for reading is 515 and the SE equals 5.8, we would expect the mean for this group of students (i.e., males) to fall between 509.2 (i.e., 515 - 5.8 = 509.2) and 520.8 (i.e., 515 + 5.8 = 520.8) 68% percent of the time[3].

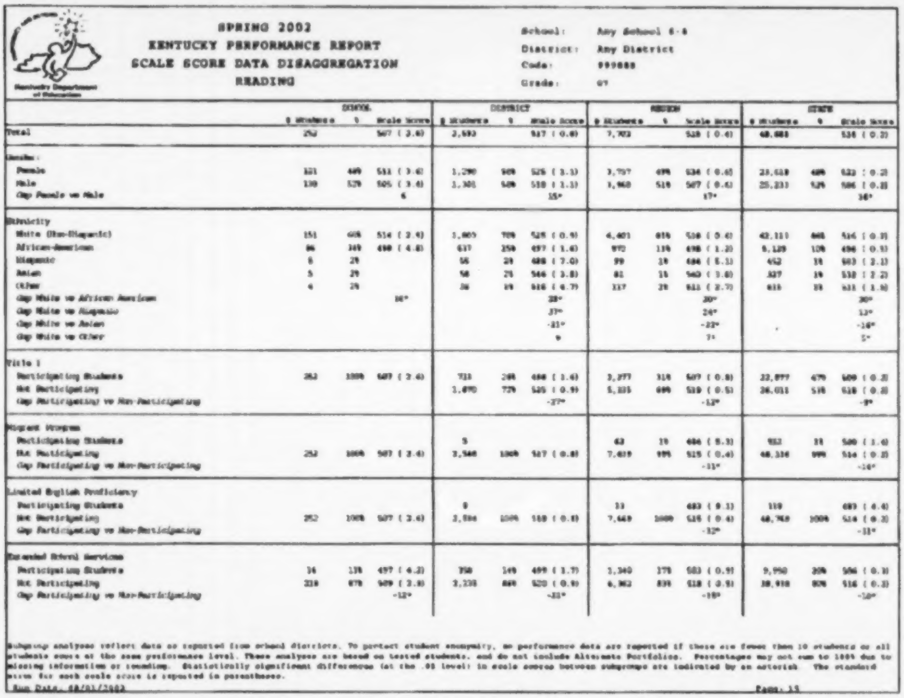
In addition to scale score means and standard errors, the *difference* or Gap between the scale score means for the following student groups are provided:

Gap between:

- ✓ Female vs. Male
- ✓ White vs. African American
- ✓ White vs. Hispanic
- ✓ White vs. Asian
- ✓ White vs. Other
- ✓ Title I: Participating vs. Non- Participating
- ✓ Migrant Program: Participating vs. Non- Participating
- ✓ Limited English Proficiency: LEP vs. Non- LEP
- ✓ Extended School Services: Participating vs. Non- Participating
- ✓ Gifted and Talented Program: Participating vs. Non- Participating
- ✓ Free and Reduced Lunch Program: Participating vs. Non- Participating
- ✓ Vocational/ Technical Education: 3 Credits vs. Non- Voc/ Tech.
- ✓ Vocational/ Technical Education: Not Concentrating vs. Non- Voc/ Tech.
- ✓ Disability Status: With vs. Without.

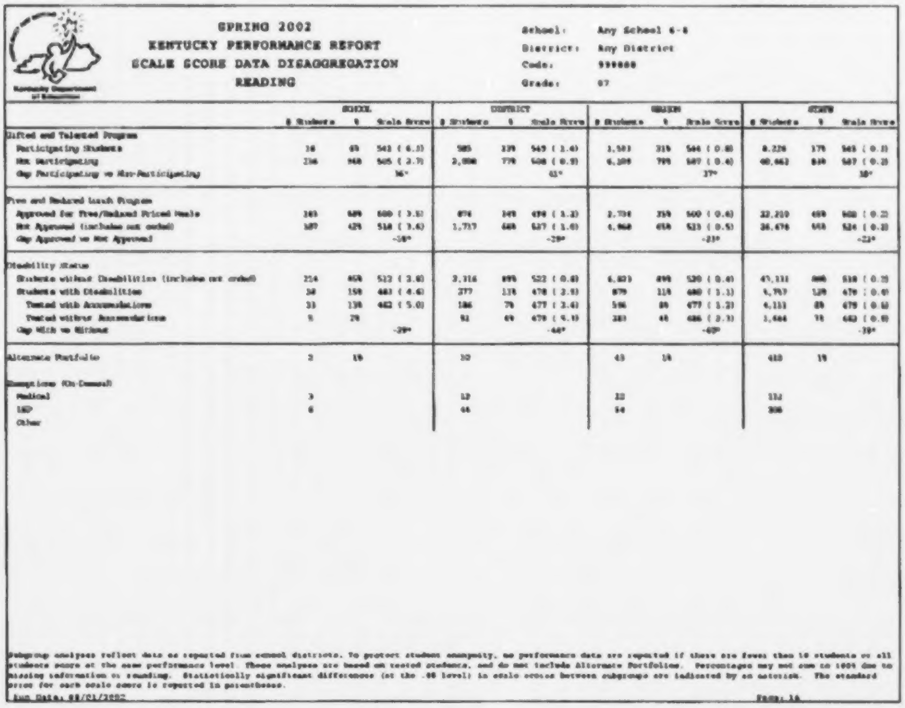
---

[3] Recall that 68% of a normal distribution falls within plus or minus one standard deviation of the mean. The SE represents an estimate of the standard deviation for the population of students on which the sample mean was calculated.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
SCALE SCORE DATA DISAGGREGATION
READING

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07



SPRING 2002
KENTUCKY PERFORMANCE REPORT
SCALE SCORE DATA DISAGGREGATION
READING

School: Any School 6-8
District: Any District
Code: 999888
Grade: 07

The Gap between the scale scores for the above student groups are reported below the mean scale score values. For example, if the mean scale scores for females and males were 507 and 515, respectively, the Gap reported would be –8 (i.e., 507 – 515 = –8). The values reported for the Gap also includes a test for statistical significance. The following formula for the standard error of the difference between uncorrelated means was used[4]:

$$\text{SEM} = \sqrt{SE_1^2 + SE_2^2} \qquad (2)$$

Where:

SEM is the standard error of the difference between two mean scores,
SE_1 is the standard error of the school mean for one student group (e.g., females), and
SE_2 is the standard error of the school mean for another student group (e.g., males).

Each value for the SEM produced by formula (2) (note that these values are not included on the report) was then multiplied by 1.96, or the Z-score used to give a two-tailed test of statistical significance at the .05 level of significance. Gap values that are statistically significant beyond the .05 level are "flagged" by an asterisk (*). These flagged values, and thus the difference between the two student groups, represents the starting point for further investigation of these differences. For example, the data disaggregation provided on the KPR can be used to further study the percentage of Novice, Apprentice, Proficient and Distinguished students for each student group. If there are no Gaps that are "flagged" by an asterisk, a general rule of thumb is to focus on Gaps or differences greater than or equal to 10 scale score points. In general, during the Standard Setting process conduced in 2001, Kentucky teachers discovered that moving a cut-point 10 or more scale score units had possible implications for the grade level, content area Descriptions of student performance, and thus our expectations of students. It should be noted that if all Gap values on these pages of the KPR were less than 10, the next strategy would be to look at Gap values relative to each other. For example, if the highest Gap values obtained for your school were around 7 or 8, then these student groups should represent the starting point for further investigation of differences. Of course, the state goal is for no, or zero gap between the performances of all student groups. As such, the state goal is that there be no gap in performance at all.

Two cautionary notes should be kept in mind when reviewing disaggregation data for schools: 1) the accuracy of the disaggregated data is dependent on how schools filled in this information on the Student Response Booklets and 2) if fewer than ten students were reported in a school or district for a category, or more than ten students scored in a category but all these students scored at the same performance level (e.g., all were Apprentice), no disaggregated data was provided to ensure the protection of the privacy of individual students. With these cautions in mind, data disaggregation information can be helpful to schools and districts in evaluating student performance in relation to special educational programs, e.g., Title 1, Extended School Services

---

[4] While it probably would have been more appropriate to use the formula for the standard error of the difference between correlated means, the more conservative formula for the difference between uncorrelated means use used. This was done in part because the test for statistical significance used in the KPR did not take into account multiple comparisons or family wise error rate.

(ESS). This information can also be used in consolidated planning to address issues relevant to equity across diverse student groups.

The Title 1 disaggregation has a few characteristics unique to the Title 1 program, which need to be noted. If a school participates in a school-wide Title 1 program, the disaggregation of student performance is for all students in the school. If a school participates in a Title 1 Targeted Assistance program, only the students participating in this program are part of the disaggregation data (as indicated on student answer document by school staff).

Some important questions school personnel may want to ask include:

- Describe any significant differences found in the school's student groups that are not found at the district, region or state levels.
- Are there any student groups at the school level where no significant differences exist?
- How does this type of disaggregation impact instructional choices and decisions?

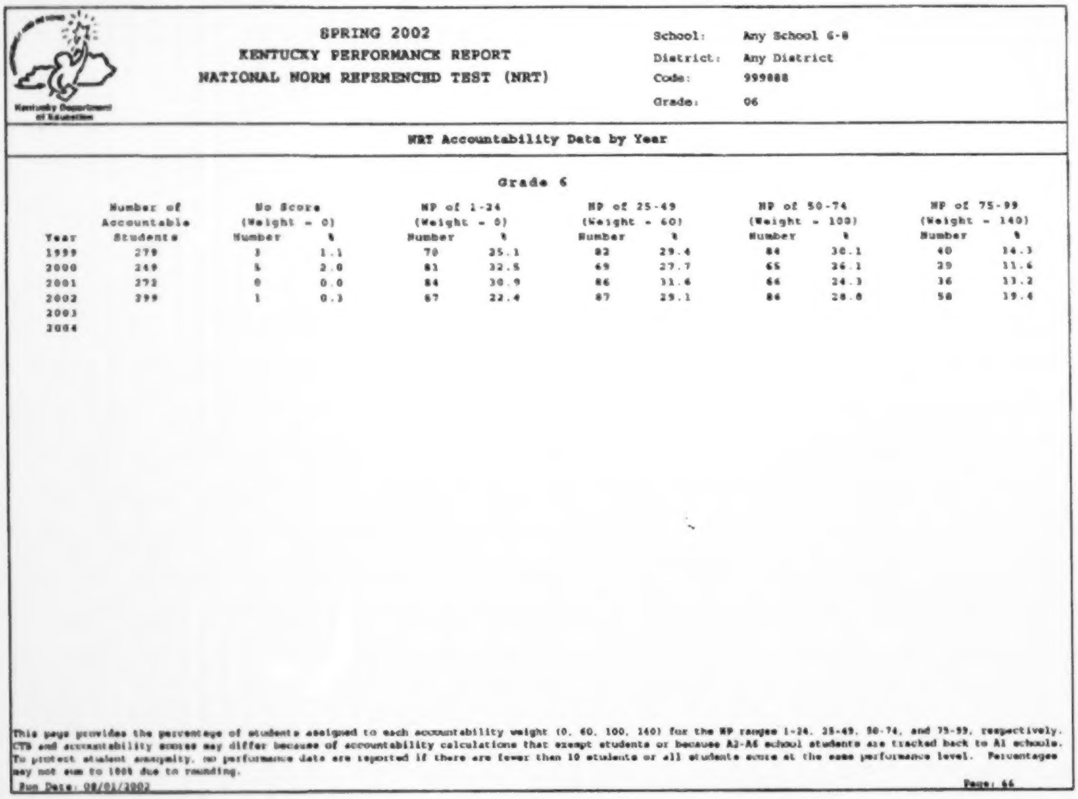
National Norm-Referenced Test (NRT)

This page follows all the KCCT content area reports and is the first of two pages providing results for the National Norm-Referenced Test or the CTBS/5 Survey. The report provides data for the NRT component of your school's accountability classification. More specifically, this page of the KPR gives the percentage of students assigned to each accountability weight (i.e., 0, 60, 100, 140) for the National Percentile ranges 1-24, 25-49, 50-74, and 75-99, respectively. State mandated components include the tests for Reading, Language and Mathematics. The NP reported is for the Total Battery composite composed of these same three tests. An example of the report is provided on the following page.

The results reported on National Norm-Referenced Test page of the KPR include only those students for which a school is held accountable. While the following four labels are used for reporting: Q1, Q2, Q3 and Q4, the percentages do *not* actually reflect the percentage of student in each Quartile. Rather, the values reflect the percentage of students scoring within an NP range defined by the Kentucky Board of Education. More specifically, the NRT component of the state's accountability system is based upon the CTBS/5 Survey (state required components) Total Battery National Percentile. The accountability "index" for the NRT is an average of student scores assigned or weighted as follows:

Weight	National Percentile
0	1 - 24
60	25 - 49
100	50 - 74
140	75 - 99

The above assignment of weights or scores puts the NRT onto the same 0 to 140 scale as the KCCT content areas. The mean index score for students on this new scale is weighted 5% in accountability. The number and percentage of students receiving each weight is given for all four years of the accountability cycle.



SPRING 2002
KENTUCKY PERFORMANCE REPORT
NATIONAL NORM REFERENCED TEST (NRT)

School: Any School 6-8
District: Any District
Code: 999888
Grade: 06

NRT Accountability Data by Year

Grade 6

Year	Number of Accountable Students	No Score (Weight - 0) Number	%	NP of 1-24 (Weight - 0) Number	%	NP of 25-49 (Weight - 60) Number	%	NP of 50-74 (Weight - 100) Number	%	NP of 75-99 (Weight - 140) Number	%
1999	279	3	1.1	70	25.1	82	29.4	84	30.1	40	14.3
2000	249	5	2.0	81	32.5	69	27.7	65	26.1	29	11.6
2001	272	0	0.0	84	30.9	86	31.6	66	24.3	36	13.2
2002	299	1	0.3	67	22.4	87	29.1	86	28.8	58	19.4
2003											
2004											

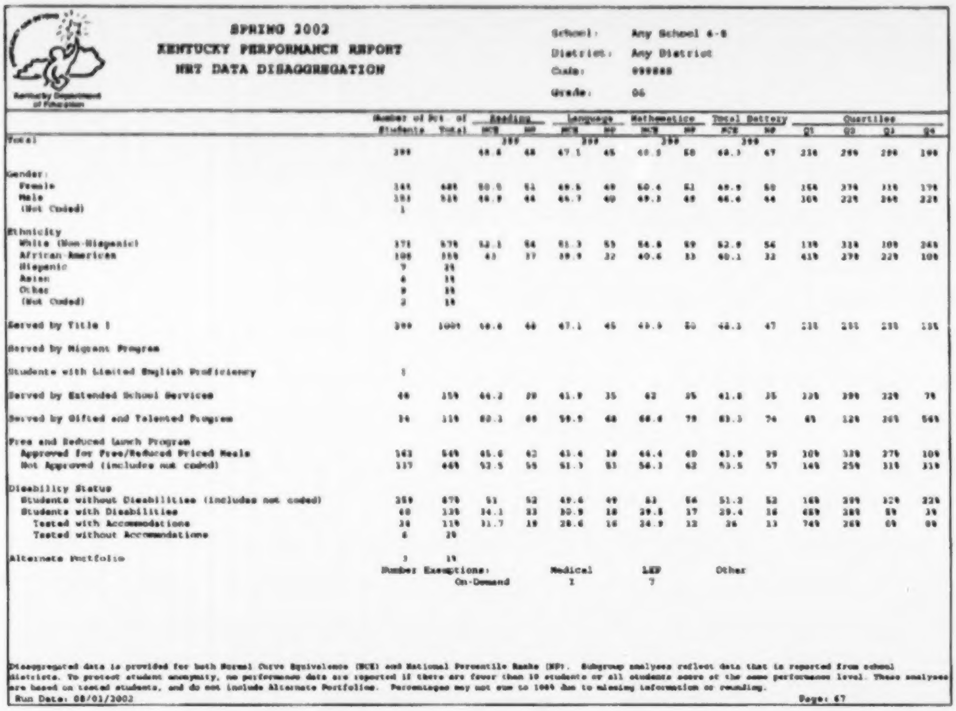
This page provides the percentage of students assigned to each accountability weight (0, 60, 100, 140) for the NP ranges 1-24, 25-49, 50-74, and 75-99, respectively. CTB and accountability scores may differ because of accountability calculations that exempt students or because A2-A6 school students are tracked back to A1 schools. To protect student anonymity, no performance data are reported if there are fewer than 10 students or all students score at the same performance level. Percentages may not sum to 100% due to rounding.

Run Date: 08/01/2002 Page: 66

NRT Data Disaggregation

For the state mandated components of the CTBS/5 Survey, important comparisons are provided for the same student groups given on other pages of the KPR. Note that the percentages on this page may not match values previously reported by CTB McGraw-Hill for the following reasons: the percentiles included in the quarters are slightly different, data excludes students exempted from accountability and data may include students that were tracked back to your school from a non-A1 school.

BEST COPY AVAILABLE



SPRING 2002
KENTUCKY PERFORMANCE REPORT
NRT DATA DISAGGREGATION

School: Any School 6-8
District: Any District
Grade: 06

An example of the NRT Data Disaggregation is provided above. As previously noted, the state mandated components include the tests for Reading, Language and Mathematics. A Total Battery composite composed of these same three tests is also reported. Note that the results reported on this page of the KPR include only those students for which a school is held accountable. In addition to the number of student tested and the percentage of total students tested, values for Normal Curve Equivalence (NCE) and National Percentiles (NP) are reported. NCEs and NPs are reported for all four scores (i.e., Reading, Language, Mathematics and Total Battery composite). The percentage of students scoring in each of the following accountability NP ranges is also provided:

Labeled	**National Percentile Range**
Q1	1 - 24
Q2	25 - 49
Q3	50 - 74
Q4	75 - 99

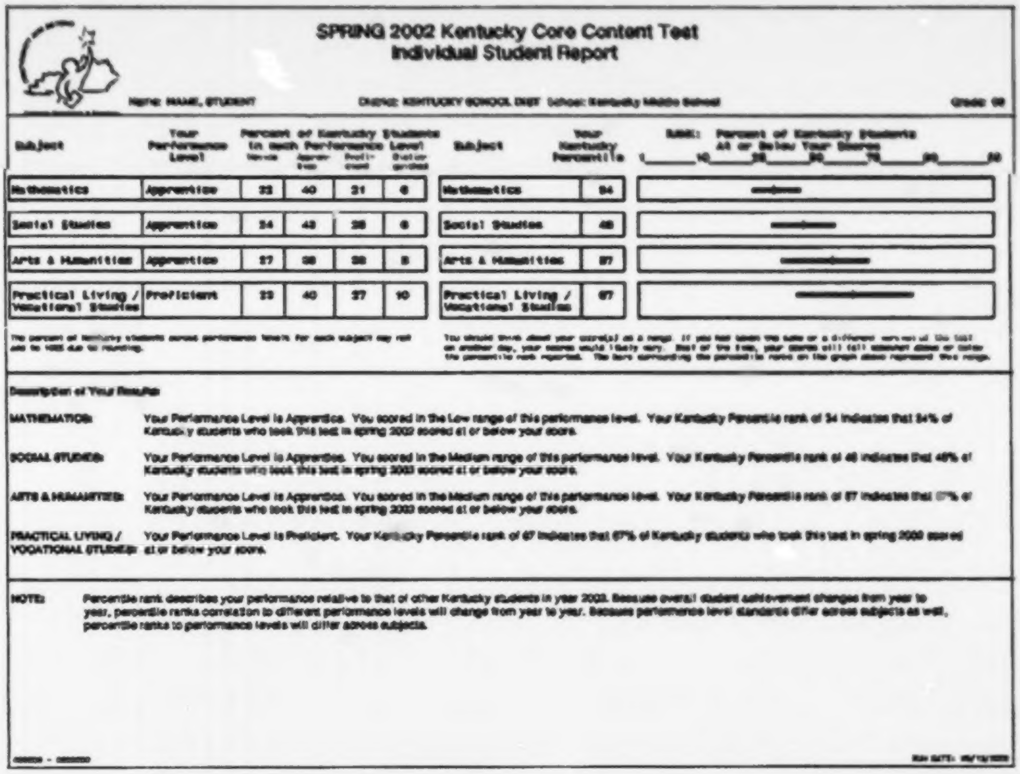
One possible use of this NRT Data Disaggregation report is to study the percentage of students scoring in Q1 through Q4 for each student group. In this way, the relative "contribution" of each student group to the NRT accountability index can be determined, thus providing guidance with respect to instructional resources and/or priorities.

Individual Student Report

The Individual Student Report (see the following page for an example) informs students and parents about individual student performance in the assessment program. Student answers to open-response questions were evaluated on a scale of 0-4, with higher scores associated with more complete and accurate responses. Multiple-choice questions were given a raw score value of 1 for a correct answer and 0 for an incorrect answer. The main feature of the report is the student's performance level (Novice non-performance, Novice medium, Novice high, Apprentice low, Apprentice medium, Apprentice high, Proficient, Distinguished), along with his/her Kentucky percentile ranking in each content area. The performance levels and percentiles are based on students' responses to both the open-response and multiple-choice questions. If a student was not tested, there will be no performance level or percentile information printed on the student reports. The Description of Results box will be marked "Non-tested" for each content area.

For students taking the same content area test during the 2001-2002 school year, the percentile rank shows where each student ranked in relation to other students throughout Kentucky. However, emphasis needs to be placed on the performance level achieved by each student. It is the performance level that determines improvement in the accountability index and determines how close a school is to bringing all students to the state goal of Proficient. Performance levels, and a clear explanation of the standards required of students, carry the most weight in CATS because they reflect the instructional strategies most valued by the state. Therefore, it is important that discussions of the reports with parents include information explaining the performance levels. As previously noted, specific descriptions by grade level and content area can be found on KDE's website at http://www.kde.state.ky.us/. In addition to this resource, a brief document, *CATS 2002 Information Sheet: Basic Information About Your Score Reports*, is available at the same website address. This document includes a Glossary of basic terms and may be useful when communicating with parents and other stakeholders.

To provide students, parents and schools with a better understanding of where a student stands in the Novice and Apprentice performance levels, the *text* in the Description of Results box identifies a student's performance as being either Novice non-performance, Novice medium, Novice high or Apprentice low, Apprentice medium, Apprentice high. These ranges (from non-performance/low to high) were determined by splitting the range of scores, at each of the Novice and Apprentice performance levels, approximately into thirds. This was *not* done at the Proficient and Distinguished levels because these students had met the state goal of Proficient. The "non-performance" Novice rating was assigned to students who earned a scale score of 325 (the lowest scale score possible), which generally reflects less than chance performance on the test. As in previous years, two copies of each individual student report are provided for students in grades 4, 5, 7, 8, 10 and 11. One copy is to be sent to parents/guardians; the other copy is for school records. For grade 12 students, only single copies (for school records) of the individual student reports have been provided.



SPRING 2002 Kentucky Core Content Test
Individual Student Report

Name: NAME, STUDENT　District: KENTUCKY SCHOOL DIST　School: Kentucky Middle School

Subject	Your Performance Level	Percent of Kentucky Students in each Performance Level: Novice	Apprentice	Proficient	Distinguished	Subject	Your Kentucky Percentile
Mathematics	Apprentice	33	40	21	6	Mathematics	34
Social Studies	Apprentice	24	43	28	6	Social Studies	48
Arts & Humanities	Apprentice	27	38	28	8	Arts & Humanities	37
Practical Living / Vocational Studies	Proficient	23	40	27	10	Practical Living / Vocational Studies	67

Description of Your Results

MATHEMATICS: Your Performance Level is Apprentice. You scored in the Low range of this performance level. Your Kentucky Percentile rank of 34 indicates that 34% of Kentucky students who took this test in spring 2002 scored at or below your score.

SOCIAL STUDIES: Your Performance Level is Apprentice. You scored in the Medium range of this performance level. Your Kentucky Percentile rank of 48 indicates that 48% of Kentucky students who took this test in spring 2002 scored at or below your score.

ARTS & HUMANITIES: Your Performance Level is Apprentice. You scored in the Medium range of this performance level. Your Kentucky Percentile rank of 37 indicates that 37% of Kentucky students who took this test in spring 2002 scored at or below your score.

PRACTICAL LIVING / VOCATIONAL STUDIES: Your Performance Level is Proficient. Your Kentucky Percentile rank of 67 indicates that 67% of Kentucky students who took this test in spring 2002 scored at or below your score.

NOTE: Percentile rank describes your performance relative to that of other Kentucky students in year 2002. Because overall student achievement changes from year to year, percentile ranks correlation to different performance levels will change from year to year. Because performance level standards differ across subjects as well, percentile ranks to performance levels will differ across subjects.

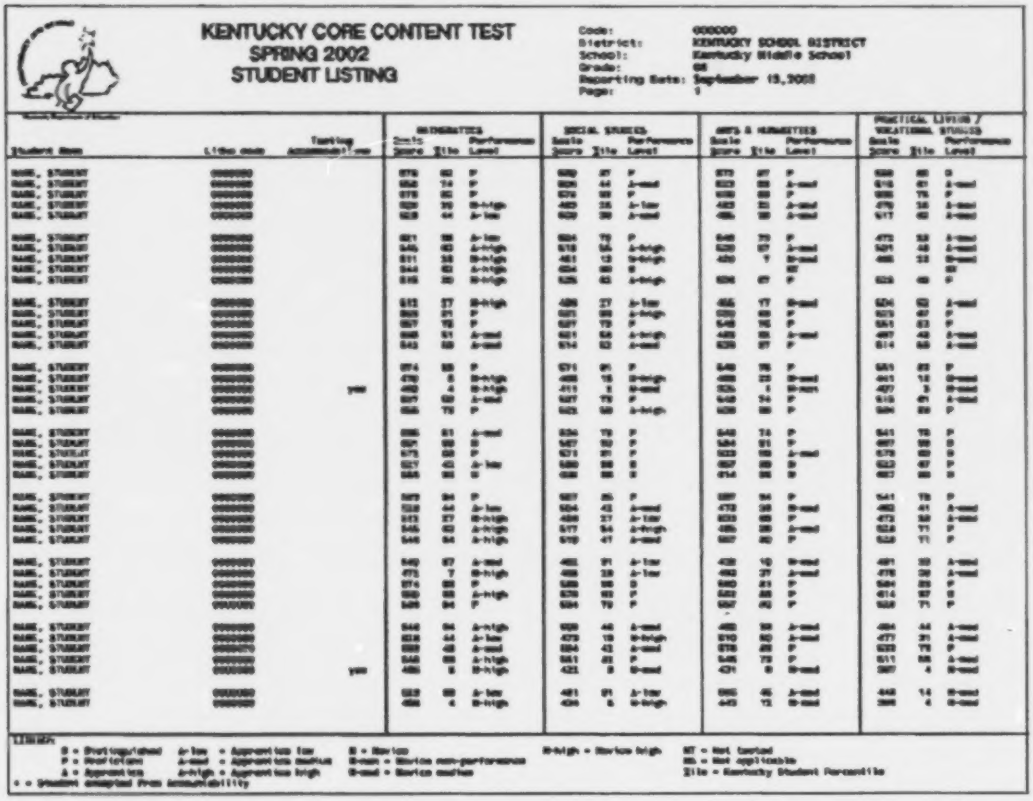
Student Listing

The Student Listing (distributed in print only on yellow paper) provides all the information in the Individual Student Report in a concise and convenient form. An example of the report is presented on the following page. For each student, the report lists the student's name, lithocode number (the student identification number for the current year of the assessment system) and performance level in the content areas of reading, mathematics, science, social studies, arts and humanities and practical living/vocational studies. The report shows the Kentucky percentile ranking in the above content areas and testing accommodations used by students when such accommodations were indicated on the Student Response Booklet.

Performance levels are based on the student's responses to the open-response and multiple-choice questions. The performance levels are abbreviated as follows:

- **D** indicates that the student scored at the Distinguished (highest) level.
- **P** indicates Proficient (the high level of achievement that state law calls for all students to attain).
- **A-high** indicates high Apprentice.
- **A-med** indicates medium Apprentice.
- **A-low** indicates low Apprentice.

- **N-high** indicates high Novice.
- **N-med** indicates medium Novice.
- **N-non** indicates Non-performance.
- **I** indicates Incomplete (this is for portfolios only). The portfolio submitted by the student was not complete. For accountability purposes, Incomplete scores are treated as non-performance.
- **B** indicates Blank (this is for portfolios and the on-demand writing prompt only). The student did not make any response to the portfolio and/or to the on-demand writing prompt. For accountability purposes, Blank scores are treated as non-performance.
- **NT** indicates Not Tested. The student did not take the Kentucky Core Contest Test and/or Writing Portfolio.
- **NA** indicates Not Applicable.
- * (asterisk) indicates a school is not accountable for the student.



KENTUCKY CORE CONTENT TEST
SPRING 2002
STUDENT LISTING

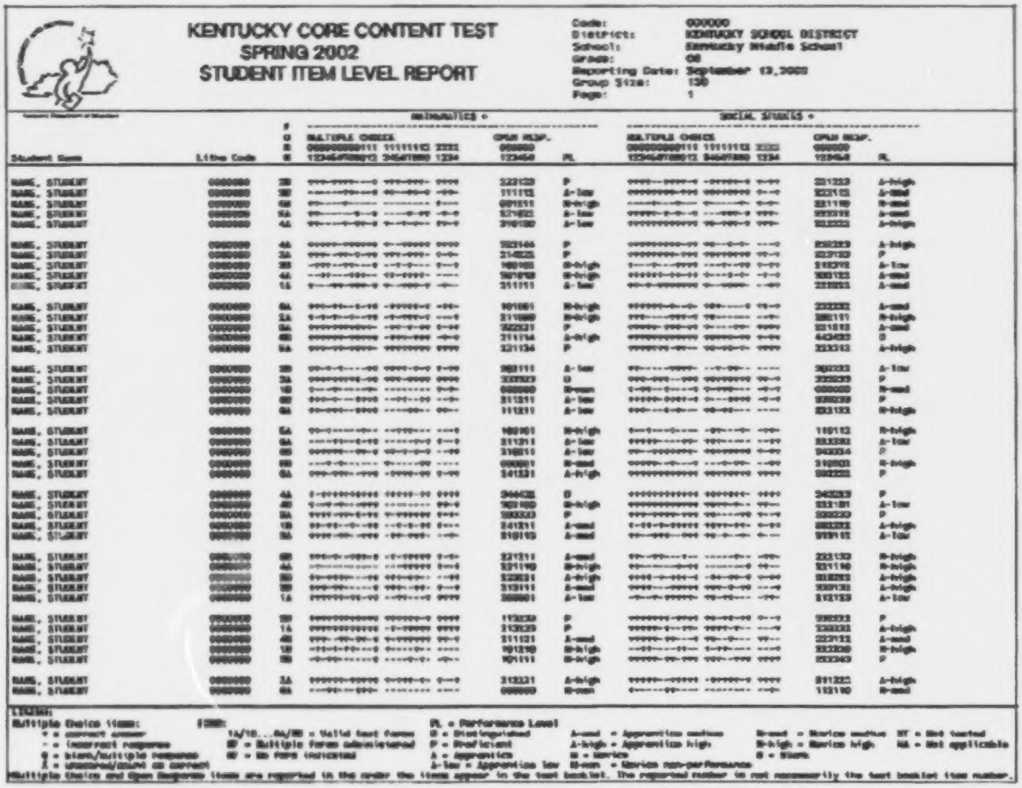
Code: 000000
District: KENTUCKY SCHOOL DISTRICT
School: Kentucky Middle School
Grade: 08
Reporting Date: September 13,2002
Page: 1

Cutpoints used to assign the four performance levels of Novice, Apprentice, Proficient and Distinguished to student work are derived from an underlying scale (see the section above on Kentucky's Accountability Index) that remains constant over time through equating. The determination of the cutpoints for non-performance, medium and high Novice is calculated by splitting the Novice interval of the scale into three approximately equal intervals. The same procedure was followed to obtain low, medium and high Apprentice performance levels. *In June 2001, the Kentucky Board of Education set new standards for the Commonwealth Accountability Testing System. The new cutpoints for determining performance levels will not vary from year to year. However, percentiles associated with the performance levels should shift reflecting student growth.*

In addition to the performance levels and percentile rankings, the Student Listing describes each student's performance in writing (Grades 4, 7 and 12). This includes a performance-level score for both the on-demand writing prompt and Writing Portfolio. Two copies of the student listing are provided, one for schools and one for districts.

Item Level Report

The Item Level Report (distributed in print only on blue paper) gives each student's score for each question on the Kentucky Core Content Test. An example of the report is presented below. The report also provides results for the on-demand writing prompt in grades 4 and 7, including each student's writing task number and score. (A table that summarizes the grades and content areas tested, including the number of open-response and multiple-choice questions asked on each of six (6) forms of the Kentucky Core Content Test, can be view in the Kentucky Core Content Test section on page 12 of this Interpretive Guide.)



KENTUCKY CORE CONTENT TEST
SPRING 2002
STUDENT ITEM LEVEL REPORT

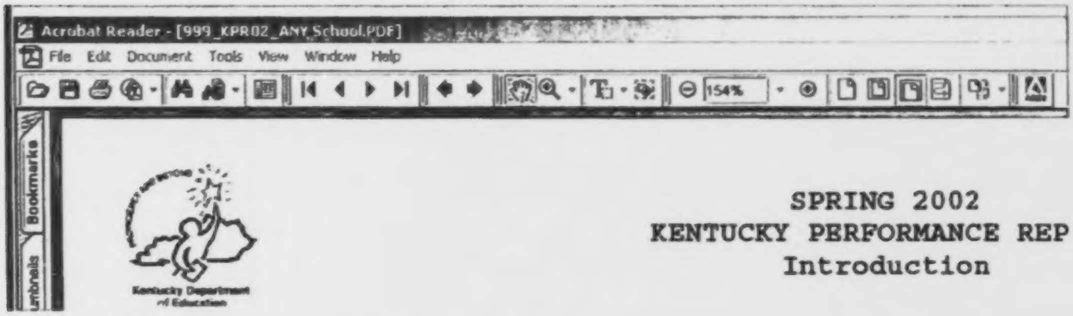
Code: 000000
District: KENTUCKY SCHOOL DISTRICT
School: Kentucky Middle School
Grade: 08
Reporting Date: September 13,2002
Group Size: 130
Page: 1

The results for the open-response items reflect how students scored on the 0-4 scale for each item. The multiple-choice items are displayed as correct, incorrect or blank. Note that the question numbers for the items on the report are only in sequential order, as such, these numbers do not necessarily reflect the actual question numbers in the form of the test taken by the student. Item Level Reports are provided for grades 4, 5, 7, 8, 10 and 11. Copies go to the school and the district.

Creating Custom Presentations Using the KPR

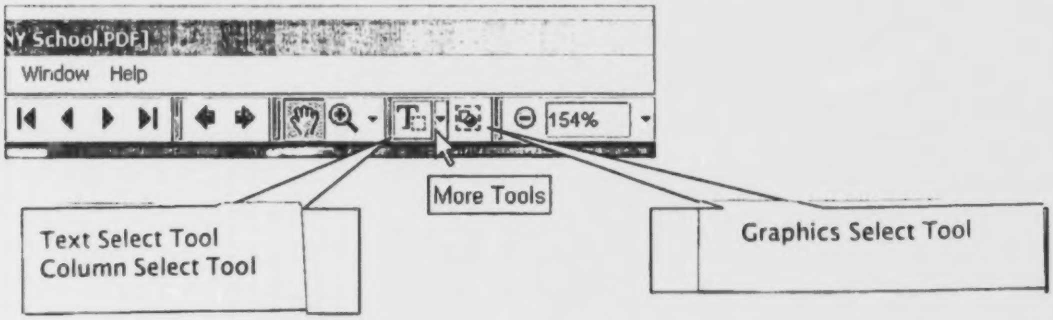
In light of the large number and variety of reports available as part of the CATS system, and more specifically the KPR, presenting *all* the available data to stakeholders during a scheduled meeting can be impractical at best, and overwhelming at worst. This section explores several alternatives to presenting the entire KPR to three important groups: School Boards, SBDMs and the lay public. However, before outlining the possible "custom" reports that may be more appropriate for these stakeholder groups, a brief introduction to a powerful tool for creating these reports is discussed first.

Acrobat Instructions

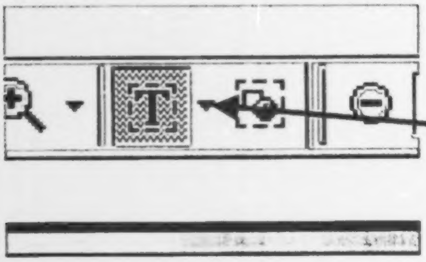

Acrobat Reader - [999_KPR02_ANY School.PDF]
File Edit Document Tools View Window Help
154%
Bookmarks
Thumbnails
Kentucky Department of Education

SPRING 2002
KENTUCKY PERFORMANCE REP
Introduction

Above is the **Menu Bar** for Acrobat Reader 5.05. Be certain that you are using Acrobat 5.0 or 5.05 because the earlier menu bar was different and the tools for cutting and pasting were not as refined. If you have an older version of Acrobat Reader, go to the Adobe Website (http://www.adobe.com/products/acrobat/readstep2.html), download your free copy of Acrobat Reader 5.05 and install this latest version of the application.


Y School.PDF]
Window Help
154%
More Tools
Text Select Tool
Column Select Tool
Graphics Select Tool

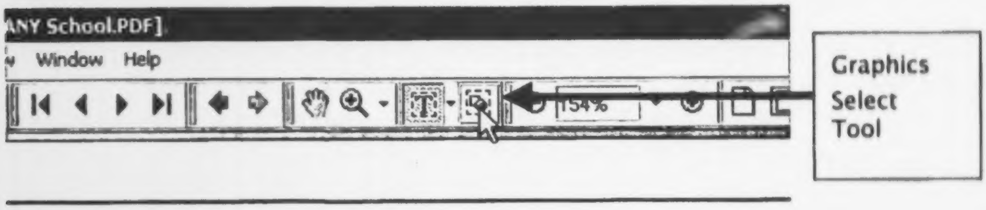
Near the center of the menu bar are the tools you will use to select text and graphics from the Kentucky Performance Reports for inclusion in your own reports in applications such as Microsoft Word, PowerPoint or Excel. To select all of the text from a page with a single text block use the **Text Select Tool**. To select portions of the text or if there are multiple columns, you must use the **Column Select Tool**. To select graphics you must use the **Graphics Select Tool**.



The close-up of the text select and graphics select tools on the left illustrates a feature of the Acrobat Reader Menu Bar. The triangular button between the two tools—when clicked—gives access to the two text tool choices. You **must** use the Column Select Tool to select text in columns for copying. *The column select tool is recommended for **all** selection and copying of text except total pages.*

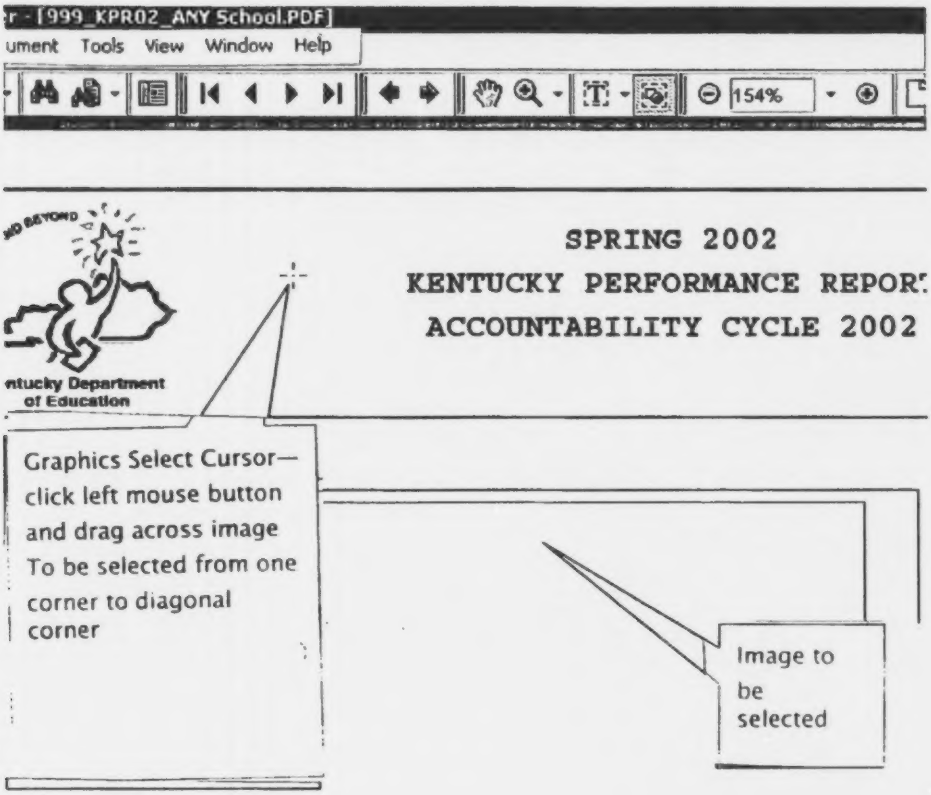
To copy a graphic from the Kentucky Performance Report (or other .PDF document) to an application such as Word or PowerPoint, do the following:

1. Left click on the **Graphics Select Tool** on the Menu Bar



SPRING 2002

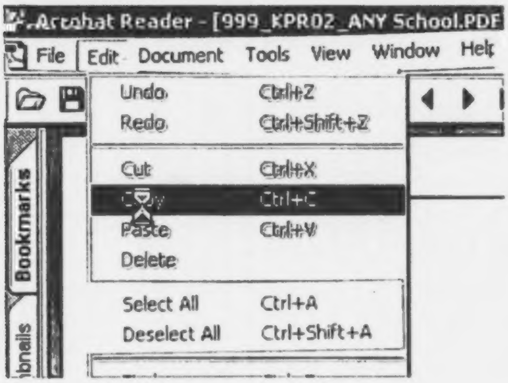
2. When the cursor changes to crosshairs, hold down the left mouse button and drag a box over the material to be selected--moving from one corner to the diagonal corner.


r - [999_KPR02_ANY School.PDF]
ument Tools View Window Help
154%
SPRING 2002
KENTUCKY PERFORMANCE REPOR
ACCOUNTABILITY CYCLE 2002
ntucky Department of Education

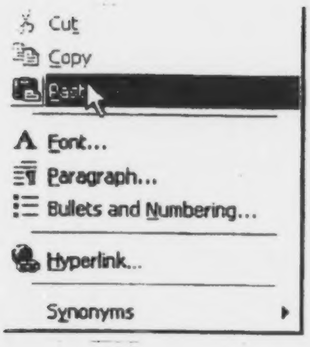
Graphics Select Cursor—click left mouse button and drag across image To be selected from one corner to diagonal corner

Image to be selected

3. When the dotted box around the material indicates that the selection is successful, left click on **Edit** on the menu bar and select **Copy** on the pull down menu.

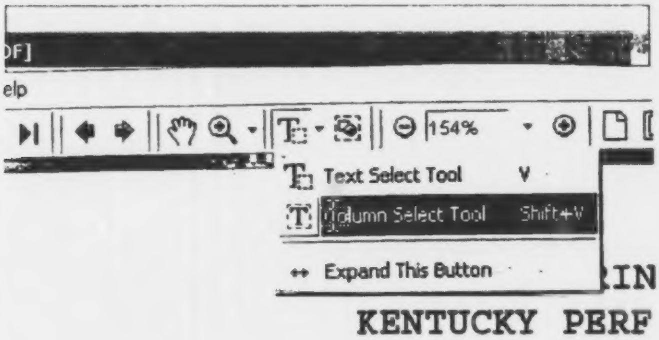

Acrobat Reader - [999_KPR02_ANY School.PDF
File Edit Document Tools View Window Help
Undo Ctrl+Z
Redo Ctrl+Shift+Z
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Delete
Select All Ctrl+A
Deselect All Ctrl+Shift+A
Bookmarks
Thumbnails

4. Open the application you wish to save the image in and locate the place you wish to insert the image, and then click on the right mouse button and select **Paste**.


Cut
Copy
Paste
Font...
Paragraph...
Bullets and Numbering...
Hyperlink...
Synonyms

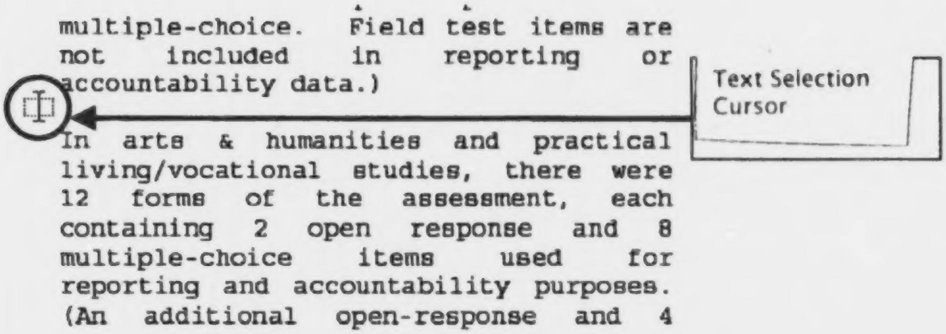
The image you have selected should now be pasted into the document you are creating in Microsoft Word or PowerPoint. (These techniques will also work in other word processing, desktop publishing or presentation applications.)

The process for selecting a block of text to copy and paste is identical, except that the **Column Select Tool** on the Menu Bar should be clicked, and when the **Column Select Tool** is being used, the cursor is not crosshairs but is a tiny dotted selection box with a bar through it (see below). The process for selecting a block of text from a multi-column layout is illustrated below:


elp
154%
Text Select Tool V
Column Select Tool Shift+V
Expand This Button
KENTUCKY PERF

1. Left click on the tiny triangle on **Menu Bar** between the Text Selection icon and the Graphics selection icon. Left click on the **Column Selection Tool** option.

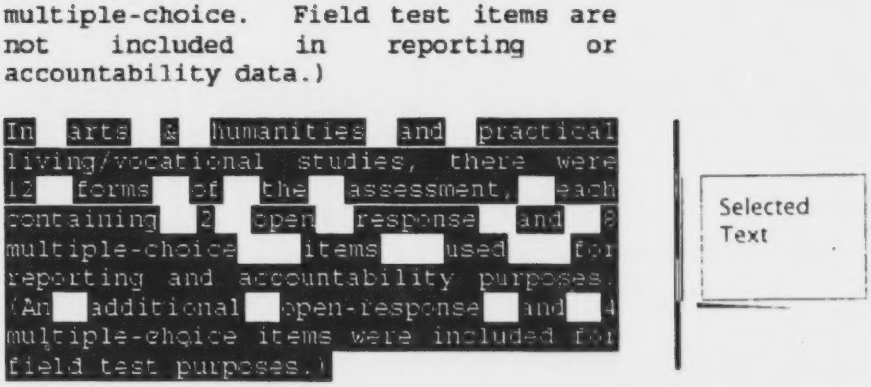
2. When the selection tool has been successfully opened, the cursor will change as illustrated:



multiple-choice. Field test items are not included in reporting or accountability data.)

In arts & humanities and practical living/vocational studies, there were 12 forms of the assessment, each containing 2 open response and 8 multiple-choice items used for reporting and accountability purposes. (An additional open-response and 4

3. Select the text block by left clicking and dragging a box over the text to be copied from one corner to the diagonal corner. When you release the mouse button the selected text will be highlighted:

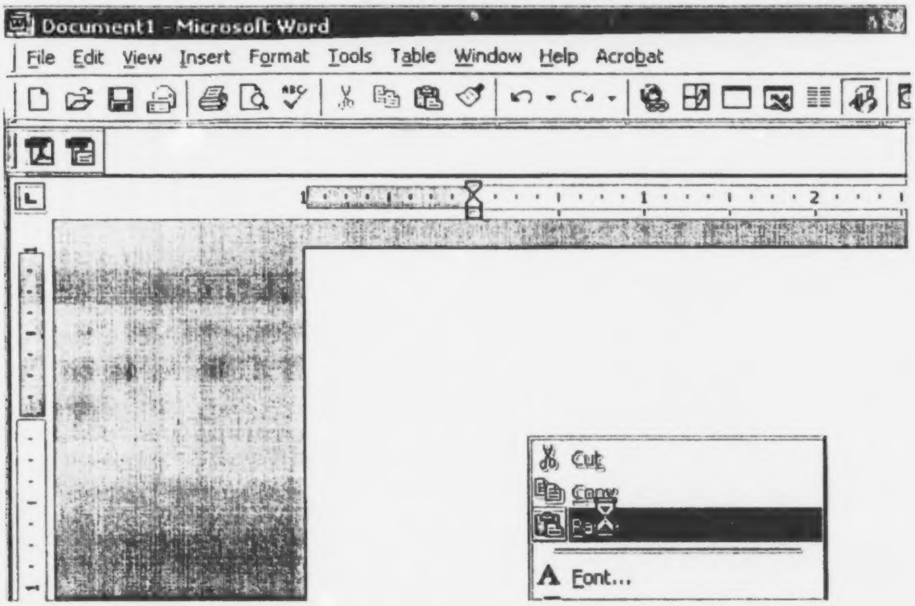


multiple-choice. Field test items are not included in reporting or accountability data.)

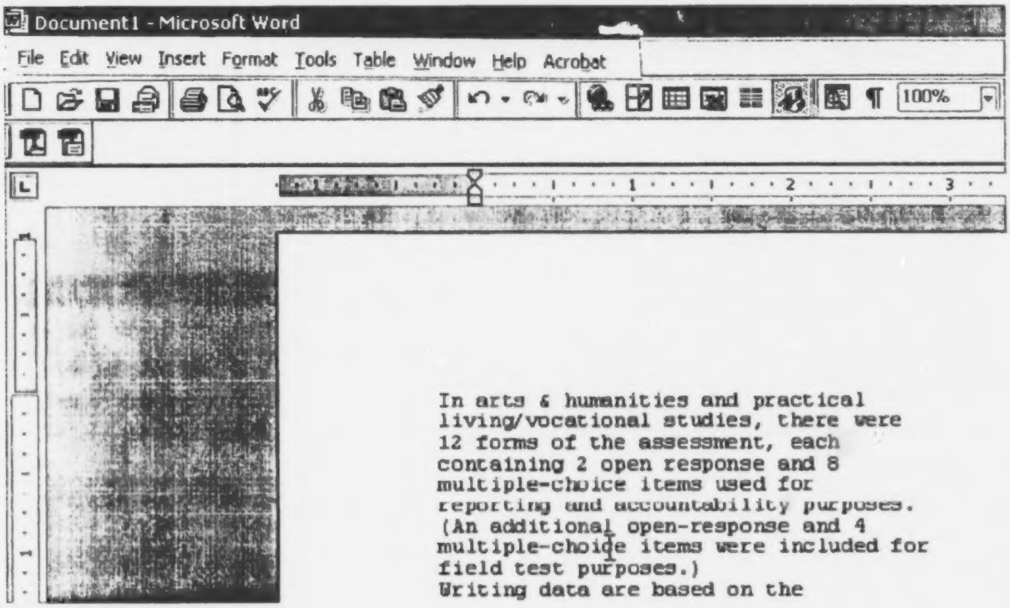
In arts & humanities and practical living/vocational studies, there were 12 forms of the assessment, each containing 2 open response and 8 multiple-choice items used for reporting and accountability purposes. (An additional open-response and 4 multiple-choice items were included for field test purposes.)

4. The copying and pasting process for text is identical to the process for graphics. When the text has been selected, left click **Edit** on the **Menu Bar**. Select **Copy** from the menu bar and left click.

5. Open the document into which you plan to paste the selected text. Click on the document where you wish to insert the text, right click and select paste:


Document1 - Microsoft Word
File Edit View Insert Format Tools Table Window Help Acrobat
Cut
Copy
Paste
Font...

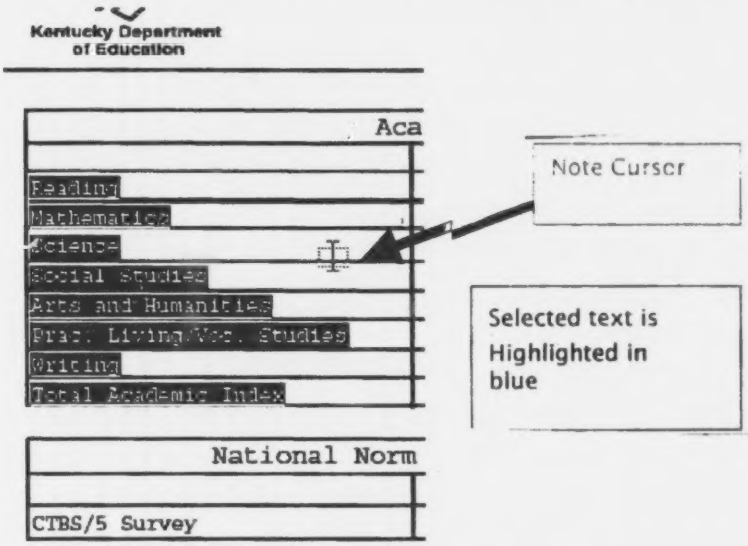
6. The pasted text will need to be reformatted, because it will be quite generic:


Document1 - Microsoft Word
File Edit View Insert Format Tools Table Window Help Acrobat
100%

In arts & humanities and practical
living/vocational studies, there were
12 forms of the assessment, each
containing 2 open response and 8
multiple-choice items used for
reporting and accountability purposes.
(An additional open-response and 4
multiple-choice items were included for
field test purposes.)
Writing data are based on the

The process for selecting information and copying and pasting into an Excel document is quite similar. **It is important that text and data be selected, copied and pasted one column at a time.**

1. Use the Column Select Tool to select the first column to be copied and pasted:

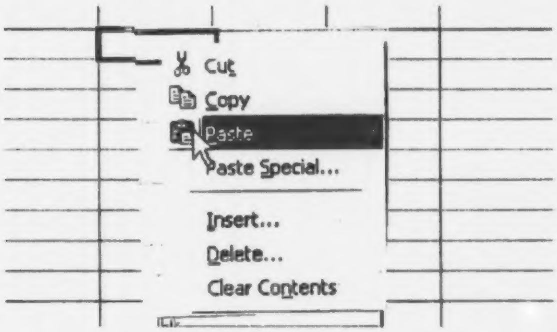

Kentucky Department of Education
Aca
Reading
Mathematics
Science
Social Studies
Arts and Humanities
Prac. Living/Voc. Studies
Writing
Total Academic Index
National Norm
CTBS/5 Survey

Note Cursor

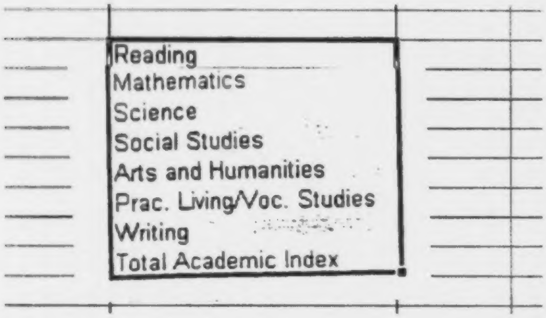
Selected text is Highlighted in blue

2. Right click on selection area and click on **Copy** or navigate to drop down **Edit** menu and select **Copy** by left clicking.

3. Navigate to the Excel workbook where you wish to paste the information from the Kentucky Performance Report. Select the top cell in your selected rectangular column and/or drag to the bottom of the column. Right click and select paste by left clicking.


Cut
Copy
Paste
Paste Special...
Insert...
Delete...
Clear Contents

4. Once pasted into the Excel worksheet, the new information can be formatted as wished.



Reading
Mathematics
Science
Social Studies
Arts and Humanities
Prac. Living/Voc. Studies
Writing
Total Academic Index

5. The selection process can continue as desired. Data, once pasted, can be manipulated and analyzed just like any other data in an Excel worksheet.

Building a Presentation for School Boards

As demonstrated in the previous section, the KPR can be "sliced-and-diced" as desired by using Acrobat Reader 5.0 or 5.05. Not only can entire pages be "lifted" from the report and put into other applications (e.g., Microsoft Word, PowerPoint or Excel), but parts or sections of individual pages can also be captured and displayed. This means that superintendents, principals, district assessment coordinators, or anyone who desires, can focus on the part of the KPR that happens to be most important for an audience and/or presentation. Along these lines, the following is but one suggested presentation for School Boards. Acrobat Reader gives you complete flexibility in the data you wish to present.

Presentations for School Boards will probably vary depending upon the size of the school district and how much time is allowed for the presentation. The following bullets summarize several options for presenting assessment results to a School Board:

- Data that presents the big picture first can be used to set the tone for a meeting. Later on in a presentation one can drill-down into the data as discussion dictates. A possible beginning point for a presentation could be a summary of the accountability results for schools in the district.

- Next, accountability trends could be summarized for elementary, middle and high schools. The Accountability Trend pages (see pages 27-28) could be copied whole, or inserted into Excel for a more custom look at the data.

- Content Area Index Trends (see pages 31-32) could be presented next by school level to note any consistent patterns among content areas as compared to the region and state.

- Next, a few examples of data from top schools in the district could be contrasted with several lower performing schools. The Trend Data, Number and Percent reports for several content areas (see pages 33-35) could be used for this contrast. The relative success and/or perceived failure of certain programs in these schools could be discuss along with the comparison data.

- In addition, comparisons with schools in other, near-by districts could also be incorporated into the presentation because the KPR for every school and district in the state is available on the Department's website.

- To address Senate Bill 168:
    - The Disaggregation Gap Trends report (see pages 29-31) could be used to look at the four-year trend for each content area.
    - Depending upon the content area(s) with the most persistent indication of a gap, the Mean Scale Scores/Standard Deviations report (see pages 43-44) could be used to further explore differences among student groups.
    - In addition, the Disaggregation, Performance Level Percents report (see pages 41-42) could be used to provide another visual presentation of the gap.
    - Finally, the Scale Score Data Disaggregation report (see pages 45-48) could be used to provide further elaboration of the Disaggregation Gap Trends report for the current year.
    - *Note that any discussion of gap data needs to include information about the number of students included in the analysis (the larger the number of students in a group, the more stable the results will be, e.g., a number based upon 50 students is better than a number based upon 10 students). This is especially true when looking at the difference between two scale score means. Also, the overlap in student group membership should be taken into account when considering the impact of certain programs. For example, targeting students with disabilities will also focus more resources on male students.*

- Finally, the possible impact of new or recommended programs could be discussed as it relates to the data presented.

Building a Presentation for SBDMs

Like School Boards, presentations for SBDMs will probably vary depending upon school size and how much time is allowed for the presentation. The following bullets summarize several options for presenting assessment results to a SBDM:

- Data that presents the big picture first can be used to set the tone for a meeting. Later on in a presentation one can drill-down into the data as discussion dictates. The beginning point for the presentation should probably be a summary of the accountability results for the school. In this case, the Accountability Cycle 2002 report with its Growth Chart, and other accountability values and targets (see pages 24-26), can be presented and explained to the counsel.

- Next, accountability trends could be summarized for the school. The Accountability Trend pages (see pages 27-28) could be copied whole, or inserted into Excel for a more custom look at the data.

- Content Area Index Trends (see pages 31-32) could be presented next to note any consistent patterns among content areas as compared to the district, region and state.

- Next, a few examples of data from other schools in the district could be contrasted with the school. The Trend Data, Number and Percent reports for several content areas (see pages 33-35) could be used for this contrast. The relative success and/or perceived failure of certain programs in the schools could be discuss along with the comparison data.

- In addition, comparisons with schools in other, near-by districts could also be incorporated into the presentation.

- To address Senate Bill 168:

    - The Disaggregation Gap Trends report (see pages 29-31) could be used to look at the four-year trend for each content area.
    - Depending upon the content area(s) with the most persistent indication of a gap, the Mean Scale Scores/Standard Deviations report (see pages 43-44) could be used to further explore differences among student groups.
    - In addition, the Disaggregation, Performance Level Percents report (see pages 41-42) could be used to provide another visual presentation of the gap.
    - Finally, the Scale Score Data Disaggregation report (see pages 45-48) could be used to provide further elaboration of the Disaggregation Gap Trends report for the current year.
    - *Note that any discussion of gap data needs to include information about the number of students included in the analysis (the larger the number of students in a group, the more stable the results will be, e.g., a number based upon 50 students is better than a number based upon 10 students). This is especially true when looking at the difference between two scale score means. Also, the overlap in student group membership should be taken into account when considering the impact of certain programs. For example, targeting students with disabilities will also focus more resources on male students.*

- Finally, the possible impact of new or recommended programs could be discussed as it relates to the data presented.

Guiding the Lay Public

Because parents, business leaders and other stakeholders in the community are not professional educators, the data presented in the KPR can be overwhelming if the proper focus is not place on the data. Appendices A and B provide two resources appropriate for those who are unfamiliar with testing in Kentucky. Appendix B contains the Glossary provided to schools for helping parents understand the Individual Student Report. Recall that these reports inform students and parents about individual student performance on the CATS assessments. Appendix C contains a 24-page document entitled, *What About Kentucky's Test?* This document uses a question and answer format to effectively convey information about general matters in testing, test construction, core content, performance standards, validity and reliability, scoring, reporting and rewards. The document provides a great beginning point for understanding Kentucky's assessment program.

More than likely, parents will be interested in their child's Individual Student Report. However, others will be interested in "How is my school doing?" For these parents, unless they are interested in the details of a specific content area, the first five reports of the KPR (i.e., Accountability Cycle 2002, Accountability Trend, Disaggregation Gap, Content Area Index Trends, Academic Index Comparisons) can be used to convey a great deal of information about how the school is doing, not only with regard to the goal of 100 in 2014, but also with respect to normative types of comparisons.

Appendix A

N/A/P/D Cut-Points

N/A/P/D CUT-POINTS IN KCCT SCALE SCORE UNITS

READING

Performance Standard Cut-Scores*

	Elem. School	Mid. School	High School
Nov Non/M	326	326	326
Nov M/H	451	428	411
NOV/APP	514	477	454
App L/M	523	488	482
App M/H	532	500	509
APP/PRO	541	511	537
PRO/DIS	601	561	584

MATHEMATICS

Performance Standard Cut-Scores*

	Elem. School	Mid. School	High School
Nov Non/M	326	326	326
Nov M/H	472	454	457
NOV/APP	548	518	523
App L/M	558	530	535
App M/H	565	543	546
APP/PRO	575	555	558
PRO/DIS	618	584	592

SCIENCE

Performance Standard Cut-Scores*

	Elem. School	Mid. School	High School
Nov Non/M	326	326	326
Nov M/H	450	434	458
NOV/APP	512	489	525
App L/M	526	498	537
App M/H	540	508	550
APP/PRO	554	517	562
PRO/DIS	588	540	608

SOCIAL STUDIES

Performance Standard Cut-Scores*

	Elem. School	Mid. School	High School
Nov Non/M	326	326	326
Nov M/H	458	430	448
NOV/APP	524	482	506
App L/M	531	499	530
App M/H	539	516	553
APP/PRO	548	533	577
PRO/DIS	586	580	621

ARTS & HUMANITIES

Performance Standard Cut-Scores*

	Elem. School	Mid. School	High School
Nov Non/ Nov	326	326	326
NOV/APP	503	478	491
APP/PRO	575	529	554
PRO/DIS	631	610	598

PRACTICAL LIVING / VOCATIONAL STUDIES

Performance Standard Cut-Scores*

	Elem. School	Mid. School	High School
Nov Non/ Nov	326	326	326
NOV/APP	460	465	458
APP/PRO	507	520	506
PRO/DIS	588	570	578

*Performance Standard levels refer to: Novice Non-Performance/Medium; Novice Medium/High; Novice/Apprentice; Apprentice Low/Medium; Apprentice Medium/High; Apprentice/Proficient; Proficient/Distinguished. Novice Non-Performance is 325 in all content areas.

Appendix B

Glossary

Spring 2002 Commonwealth Accountability Testing System
Individual Student Report

Glossary

Spring 2002 Commonwealth Accountability Testing System
Individual Student Report

Spring 2002 Commonwealth Accountability Testing System - The testing/assessment program used to test/assess the progress being made by Kentucky schools. The program is made up of five parts:

1) Kentucky Core Content Tests at grades 4, 5, 7, 8, 10, 11 and 12
2) Writing Portfolios at grades 4, 7 and 12
3) Alternate Portfolios at grades 4, 8 and last anticipated year
4) Non-academic index, which includes:
   - Attendance and retention at the elementary level.
   - Attendance, retention and dropout rates at the middle school level.
   - Attendance, retention, dropout rates and successful transition to adult life at the high school level.
5) Norm-Referenced Tests assessing reading, language arts and mathematics at the end of Primary, grades 6 and 9.

The Kentucky Core Content Test, Norm-Referenced Tests and Writing and Alternate Portfolios produce individual student information. Non-academic data components produce data only at the school and district level.

NAPD Descriptions - The following are summaries of the language used to describe Novice, Apprentice, Proficient, and Distinguished. These categories are used in reporting student results within the Commonwealth Accountability Testing System. The Proficient level is the long-term goal for all students. *For more explicit and detailed descriptions it is best to consult the descriptions for each particular grade level and content area.* These descriptions can be found on the Kentucky Department of Education's (KDE) website at http://www.kde.state.ky.us/.

Novice	* Student demonstrates minimal, limited, underdeveloped, and at times inaccurate content knowledge and reasoning. * Student communication is ineffective and lacks detail with no evidence of connections within or between content areas. * Student uses strategies that are inappropriate.
Apprentice	* Student demonstrates some basic content knowledge and reasoning ability. * Student communicates reasonably well but draws weak conclusions or only partially solves or describes. * Student attempts appropriate strategies with limited success.

Proficient * Student demonstrates broad content knowledge and is able to apply it.
* Student communication is accurate, clear, and organized with relevant details and evidence.
* Student uses appropriate strategies to solve problems and make decisions.
* Student demonstrates effective use of critical thinking skills.

Distinguished * Student demonstrates an in-depth, extensive, or comprehensive knowledge of content.
* Student communication is complex, concise, and sophisticated with thorough support, explicit examples, evaluations, and justifications.
* Student uses and consistently implements a variety of appropriate strategies.
* Student demonstrates insightful connections and reasoning.

To communicate a more specific indication of how close a student's work is to the next performance level, for reporting purposes in reading, mathematics, science and social studies, the Performance Levels of Novice and Apprentice are subdivided into the following categories:

- Novice Non-performance
- Novice Medium
- Novice High
- Apprentice Low
- Apprentice Medium
- Apprentice High

Performance Levels are derived for the Kentucky Core Content Test by taking a weighted sum of the performances on open-response and multiple-choice items and converting it to an appropriate Performance Level. Performance levels are derived from student Writing Portfolios through a process of training local school staff to apply the scoring standards to the portfolio as a whole in a consistent manner. Alternate Portfolios are scored at the regional level by trained teachers from neighboring districts.

Scoring Guides - These are guides that are used to score student answers. For open-response questions, a different guide is developed for each question. Additional guides are developed for Writing Portfolios and Alternate Portfolios.

Kentucky Core Content Test - This is the test taken by students in grades 4, 5, 7, 8, 10,11 and 12 in the spring of the school year. At grades 4 and 7, this test contains open-response (essay-like) and multiple-choice questions in reading and science. It also has two writing questions (prompts); students select and write a response to one of those prompts. At grades 5 and 8 the test contains open-response and multiple-choice questions in mathematics, social studies, arts & humanities and practical living/vocational studies. At grade 10 the test contains open-response and multiple-choice questions in reading and practical living/vocational studies. At grade 11 the test contains open-response and multiple-choice questions in mathematics, science, social studies and arts & humanities. At grade 12 the test has two writing questions (prompts); students select and write a response to one of those prompts.

Portfolios - These are collections of each student's best work. Writing and Alternate Portfolios are developed over time as part of the accountability program in the following grades:

Writing Portfolios	grades 4, 7 and 12
Alternate Portfolios	grades 4, 8 and last anticipated year

The Alternate Portfolio refers to a measurement process used with students generally thought to have severe disabilities and who are not able to participate within the normal curriculum, even when they are provided all possible accommodations and adaptive devices available. This portfolio program typically involves less than 1% of the total student population.

Kentucky Percentile Rank - This number describes how a student performed on the test compared to other Kentucky students who took the same test in the same year. For example, if a fourth grade student's Kentucky Percentile Rank in reading is 53, 53% of the Kentucky fourth grade students who took the reading test in the same year scored lower than or equal to the student.

Standard Error of Measurement – One way to think about the standard error of measurement is to think about a test score as being a single score contained within a range of other possible scores. For example, if you had taken the same test or a different version of the test on another day, your scores would likely vary. Most of the time your scores would fall within several percentile points of your true abilities. If it were possible to re-test a student on the same or a different test numerous times, the student would usually score within a band of scores defined by the current score plus/minus *one* standard error of measurement. If one were to consider a score range defined by the current score plus/minus *one* standard error of measure, the student would score within this range approximately 65% of the time. The score range gives a more complete picture of a student's score possibilities. Educators know this, and in fact, specifically ask that score ranges be included with scores. The *standard error of measurement* is a standardized statistic used by test developers to indicate the measurement accuracy of an assessment. Standard errors of measurement are used with the Kentucky Core Content Test, as well as many other tests, including tests like the ACT and SAT.

Score Range *(Graphically displayed around student Kentucky Percentile Ranks)* - On the Individual Student Reports, a student's Kentucky Percentile Rank is graphed as a point surrounded by a bar. The point is the Kentucky Percentile Rank. The bar is the *score range*. The point and the bars represent the student's score plus/minus one *standard error of measurement* (see definition above). The bars around a student's score in each subject show the range of scores the student would likely have received if he/she had taken the same test, or a different version of the test, on another day. It should be noted that all tests contain measurement error for a variety of reasons, including environmental factors (e.g., testing conditions) and student factors (e.g., fatigue, stress). Because of this, any student level score should be interpreted as representing a range of possible scores, or a *score range*.

Appendix C

What About Kentucky's Test?

WHAT ABOUT KENTUCKY'S TEST?

THE QUESTIONS

Kentucky parents and teachers ask a lot of questions about Kentucky's testing program. Some of the basic questions are: What does it cover? How is the test built? Who builds it? Is it valid? What is a standards-based test? Why does it take so long to get the results? Can an essay really be graded consistently? Why aren't students held accountable? These are just a few of the questions asked that we will consider.

GENERAL MATTERS

WHAT IS TESTED AND WHEN IS IT TESTED?

The first time a student meets the Kentucky testing system, which is called the Commonwealth Accountability and Testing System (CATS for short), is in the third grade. In April third graders take a multiple-choice test called the Comprehensive Test of Basic Skills (CTBS/5), which is produced by the CTB McGraw-Hill Corporation. Because this test is used nationwide, Kentucky students can be compared to students in other states. This test is repeated in grades six and nine.

In grade four, students write parts of the Kentucky Core Content Test (KCCT for short) for the first time. This test is very different from the CTBS/5. First, the students write essay type answers (called open-response), as well as multiple-choice. The open-response answers are limited to one page. A second difference is that the KCCT is designed to cover the breadth of the Core Content, which is specifically what Kentucky students are expected to know and do at the fourth grade level (more about that later). The test asks questions about reading, writing and science. For reading and science there are six open-response questions that count for the student, and one that is being evaluated for future tests. This one does not count in the student score. The writing test offers the student two questions, but they only have to answer one. In addition, fourth graders produce a collection of expanded work, representing their best efforts, called a Writing Portfolio (more about that later, too).

Grade five students continue the KCCT, but in different subjects than fourth grade. While mathematics and social studies are tested in grade five, the format of this test resembles the fourth grade reading and science test with six "live" open-response questions and one experimental question. Two new subjects are also tested for the first time: arts & humanities, and practical living/vocational studies. These tests are shorter, having two open-response items and one experimental item, along with fewer multiple-choice than what is on the mathematics or social studies part. The following bullets summarize testing in middle and high school, which is very parallel to the testing in elementary schools mentioned above:

- Students in the sixth grade take a grade appropriate version of the CTBS/5.
- Seventh graders write tests in the same subjects as fourth grade, with the same number of questions at a grade appropriate difficulty level.
- Eighth grade repeats the same subjects as fifth grade.
- Ninth grade students take a grade appropriate version of the CTBS/5.
- Tenth grade takes the KCCT in reading and practical living/vocational studies. These tests have the same number of questions as these subjects had in earlier grades, but the questions have increased in difficulty at each level.
- Eleventh grade is the most heavily tested grade in high school. Students write the KCCT in mathematics, science, social studies, and arts & humanities.
- Since many students graduate at the end of the first semester of grade twelve, only two parts of the CATS are completed: in twelfth grade, the writing portfolio which can be finished the first semester, although it is not due until April, and the writing question (called writing on-demand) which is also administered in April.

One of the strong points of CATS is that it does not depend on a single type of testing. The KCCT includes multiple-choice in every subject and grade from three through eleven, open-response in grades four, five, seven, eight, ten, and eleven, writing questions and portfolios in four, seven and twelve. The variety of testing methods allows students to show a greater range of their abilities.

HOW DOES KENTUCKY'S TESTING SYSTEM COMPARE WITH OTHER STATES?

Kentucky has an advantage over some of the largest states where the cost of scoring open-response questions forces them to use all multiple-choice, or only a few open-response in a limited number of subjects. Kentucky has an advantage over some of the smallest states, which have inadequate resources to construct a test that meets their own needs expressed in something like our Core Content.

Another advantage of CATS is that Kentucky tests more subjects than the states that limit testing to reading and mathematics, and sometimes writing. In addition to testing a larger number of subjects, Kentucky tests at as many or more grade levels than some other states. Some writers have pointed out that states that only give a single multiple-choice test in reading and math, but give it every year, have a better picture of how the individual student is keeping up nationally. In a limited sense this is true, but the test may not match very well what is taught in the state, and the breadth and richness of a curriculum like Kentucky's is lost. Such tests are much more susceptible to cheating and teaching to the test rather than teaching to a body of knowledge.

There are also differences at the high school level between Kentucky and some of the other states. Some states test at the end of specific courses, such as algebra, or U.S. history, or American literature, instead of the general mathematics or social studies tests that Kentucky uses. These "exit" exams are used for a variety of purposes from assigning final grades to entrance to a following course. The goal is to make the student more responsible, but a one-day test may not be as descriptive of a student as a semester or year of daily work.

Some states use an exam to decide whether a student is ready to graduate. Kentucky's system is devoted to improving instruction, not to testing individual students. This is an important enough issue that we will consider it further.

WHY DOESN'T KENTUCKY USE TESTS FOR PROMOTION AND GRADUATION?

Some states have what are called promotion or "exit" exams. Logically, it seems that this would put more pressure on students to do their best in order to pass to the next grade. They would be more "accountable." If the world were simple and completely logical maybe this would work, but in the real world there are some surprises hidden in such testing. For example:

- Promotion tests almost always increase the number of students held back (retained) in the prior grade, resulting in increased costs for a few years until new school population patterns are established.
- A second result is that the increased retentions lead to increased dropouts at grades eight through twelve, especially the grade prior to the administration of the test. This is the opposite of what Kentucky has been seeking to do with regard to dropouts.
- Another problem is that if the test is modified to keep approximately the same pass rate, then it does not seem to measure as much or require as much educational achievement.

So, in order to have high pass rates and a hard test, just have the teachers teach better. This brings us back to the beginning. Retaining students, according to many studies, does not motivate students to perform better, but better teaching and smaller classrooms at the primary grades do.

Kentucky's testing system aims to make very clear what is to be taught, and how good performance should look, so that teachers know what will be tested and at what difficulty level they must teach the content. Student motivation issues tend to be reduced in well-taught classrooms, but there will always be the one student in twenty who cannot be externally motivated, even by a high stakes test.

Kentucky has chosen to not ask of a standardized test something that it cannot do: that is, give a picture of the development of a child from a one-day paper and pencil test. Kentucky has sought to put data together at a level where it can give an accurate picture, which is at the school level.

WHY DOESN'T KENTUCKY HOLD STUDENTS ACCOUNTABLE?

This was partially discussed above under the promotion/graduation questions. Kentucky has considered a number of proposals to increase student accountability. The problem is strongest at the high school level, begins to appear at the middle level, and is less of an issue at the elementary level. Including KCCT performance as a small part of the student GPA is an example of a proposal that was considered and not accepted. Since the proposal was optional, most high schools indicated unwillingness to engage in the extra calculations this would involve. Other proposals have surfaced, both locally and nationally, but no really effective means of motivating low performing students, other than the classroom teacher, has been found. Kentucky will not adopt student accountability until a successful method has been found.

TEST CONSTRUCTION

WHO WRITES THE QUESTIONS FOR THE KENTUCKY TEST?

Unlike standardized tests, which may be built in other states like New Jersey or California, and may or may not be related closely to state standards, the Kentucky test is related to nationally recognized standards as well as Kentucky standards. With the exception of the CTBS/5 component, Kentuckians create the Kentucky test. The writers are Kentucky teachers who are experts in the subject area for which they write questions. They are among the best Kentucky teachers who have exhibited expertise in teaching, have shown the ability to teach by various methods to meet the wide range of student needs, and have come to the attention of their principal and/or District Assessment Coordinator who recommend them. They are assigned to a Content Advisory Committee (CAC), which meets in the spring of the year to write questions.

The CAC may write as many as thirty open-response questions, and as many as 100 multiple-choice questions in each subject area. The CAC participants write with three objectives in mind: to improve the quality of questions used on the test, to make sure all parts of the Core Content are covered, and to provide replacements for questions. Approximately 20% of the questions are replaced each year.

The questions are then submitted to one of our contractors, currently WestEd, which is a California company with expertise in question writing and building tests. They edit the questions and balance the wrong answers (the distractors) so they are not correct, but not so ridiculous that no student would choose them. In the fall of the year the CAC comes together again and looks at the revised questions and makes selections for testing. As you can see, this means that Kentucky teachers write the questions for the test and pick out which ones will be used each year.

WHO PUTS KENTUCKY'S TEST TOGETHER?

The contractor builds six different forms of the test in each subject area. The multiple forms allow the full coverage of the Core Content in that subject, which is important for evaluating a school. Each form has one experimental question that the student answers, but the form is labeled A or B which allows the testing of two experimental questions. The live questions remain the same on both forms. Several statistical measures of the quality of the question are accumulated as it is tested and used, such as percentages of students who select each answer on multiple-choice questions, p-values (a measure of difficulty), bi-serial correlations, and others. More information about these statistical matters is available in the *KCCT 2000 Technical Report* which is available from the Office of Assessment and Accountability, Kentucky Department of Education, Frankfort, Kentucky and on the website for the department. These statistical tools are also used to make the forms as comparable in difficulty as possible.

One form of the test is selected for use with visually impaired students. That form will be translated into Braille, produced in large print, and recorded on tape so a student can play it, and back it up, to hear the questions a second or third time if necessary. Some impaired students may answer on a computer. The intention is to give the impaired student the same chance that other students have to answer the questions successfully.

Another form is selected for "scaling" and "linking" the test from year to year. That form is held stable from one year to the next so that changes in performance can be measured as real changes. When the other forms are "scaled" within the year to the linking form from year to year, the gains exhibited are genuine. A pattern for selecting the linking form has been developed so that one form is not held stable for several years, which leads to "aging," or the questions becoming familiar and known to teachers, which would distort the results.

WHAT GUIDES THE CONTENT OF KENTUCKY'S TEST?

The Core Content has already been mentioned several times. What people usually mean by a standardized (norm-referenced) test is one that is connected to the expected performance of a normally distributed group of students at a particular level in school. A representative sample of students set what is presumed to be normal performance on the test. The sample sets the national mean, quartiles and percentiles, which is the way scores are usually reported.

A standards-based test is tied to a set of statements about what students should know and activities they should be able to do. The statements are fixed, and the distribution of the students will not follow a normal curve in most cases. The boundaries between categories are called cut-points. No matter how many students move into a higher category, the boundaries or standards do not change. Kentucky divides students into four categories: novice, apprentice, proficient and distinguished. The student who does nothing is categorized as novice non-performing. The goal is for all schools to be proficient by 2014. Proficient is defined as a score of 100 on a 140-point scale. For the school to achieve that goal nearly all students must also move to proficient.

The *Core Content for Assessment* is a document that states the minimum that students must know and do in terms of what will be tested. This document is available on the Kentucky Department of Education website. Students learn much in school that cannot be tested, but whatever teachers choose to teach must include the Core Content in their subject area. If the course is Algebra, many concepts will be taught that cannot be included on the KCCT, but certainly the teachers must make sure that students learn the particular algebraic concepts that are mentioned in the Core Content, because they will be tested.

JUST WHAT IS IN THE CORE CONTENT?

The Core Content describes what to know and do at three levels: elementary middle and high school. Seven subjects are included in the Core Content: reading, mathematics, science, social studies, arts & humanities, practical living/vocational studies, and writing. Each content subject is divided into subdomains: mathematics, for example, has four, which include number/computation, geometry/measurement, probability/statistics, and algebraic Ideas. Science has three subdomains: physical science, earth and space science, and life science. Other subject areas are similarly organized.

The next division of the content is that each subdomain is divided into sections. For example, the 4th grade science subdomain of earth and space science is sectioned into properties of earth materials, objects in the sky, and changes to earth and sky. The final layer in the content is the specific statement of the content under the subdomain and section. These are called "bullets." One bullet under properties of earth materials for the 4th grade says, "Earth materials include solid rocks, and soils, water and the gases of the atmosphere. Minerals that make up rocks have properties of color, texture, and hardness. Soils have properties of color, texture, the capacity to retain water, and the ability to support plant growth. Water on Earth and in the atmosphere can be a solid, liquid or gas." The teacher is told the broad topics to teach, but not how to teach it.

WHO CREATED THE CORE CONTENT?

Once again, Kentucky teachers, the experts in their fields, wrote the Core Content. The committees of teachers who did this task consulted and considered what national organizations had published. For example, the National Council of Teachers of Mathematics has extensively documented content at each grade level in ten strands.

These national content standards were considered in Kentucky's Core Content writing. Similar standards exist in language arts, science and social studies. The teachers in arts and humanities, and practical living/vocational studies had less guidance from national organizations.

HAS THE CORE CONTENT NARROWED OR DUMBED DOWN WHAT STUDENTS HAVE TO KNOW?

One of the most common accusations leveled at any statewide testing program is that teachers teach to the test and dumb down the curriculum. There are several levels of teaching to a test. The first is obtaining the test questions and drilling students over correct answers to the test. This clearly is cheating, artificially inflates student scores, and contributes very little to student learning. ,Kentucky seeks to avoid this kind of teaching by keeping the questions secure, having teachers sign non-disclosure statements, making it inappropriate to copy down the test questions, or even making a list of topics covered. This is actually not necessary since the Core Content is the list, and the six forms cover all or nearly all in a given year.

At another level, however, Kentucky does encourage teaching to the test. Since the test and the Core Content match so closely, every bit of the Core Content needs to be taught, sometimes in multiple ways. This procedure assures that students can answer whatever question comes up on that topic. In some years, questions that have been used on previous CATS tests, and that will not be used again, are released so that teachers have examples of what students have to do to succeed. Examples of student papers at the four performance levels are also released (without names of course). On the other hand, classroom topics need not be limited to the Core Content, however, one of the primary reasons schools are not successful on the CATS is that they do not teach the Core Content. This failure to address the Core Content has been revealed by the school auditing process, which has been conducted in recent years. The most successful schools have rich and varied curricula, but do thoroughly cover the Core Content.

WHAT ARE STANDARDS?

There are several kinds of standards. The Core Content, already mentioned, is one type of standard, a content standard. Every child is supposed to be able to know what the Core Content specifies, and do the skills described at the appropriate grade level of difficulty. The content standard that Kentucky uses is certainly not everything a person should know, but it is the minimum that a person must know to be considered educated and able to function in society.

A second kind of standard is a performance standard. This is a boundary mark that is the target for a student to achieve in order to be classified a certain way. In the high jump, for example, a jumper in high school who exceeds six feet six inches would be considered proficient, in college it would take a jump of six feet ten inches to be considered proficient, and a world class jumper might have to reach seven feet and a few inches to be considered proficient. These are benchmarks that indicate whether the person is going to be competitive. In education the concept is the similar. There are certain scores on a test that are benchmarks. They are called cut scores. Everyone who reaches the first cut score in Kentucky is considered an apprentice. Those below that first mark are novice. Those above the second cut score are considered proficient, and those beyond the third are considered distinguished.

WHAT IS "STANDARD SETTING?"

Standard Setting is the process of deciding where the boundaries are between the four categories that Kentucky uses in describing student accomplishment. Standards were set in 1992 for the old KIRIS test by a relatively small group of teachers. While those standards generally worked well, there were problems in some subjects in that it was difficult for students to actually show the higher performance categories. When the KIRIS was revised into CATS in 1998, it was clearly necessary to set new standards for the new test. This was done during the timeframe from late 1999 to early 2001.

Approximately 1600 teachers participated in a six-step process designed by the Kentucky Department of Education (KDE) and a panel of six national testing experts. Three different methods of setting standards were used, two of which did not even exist when Kentucky first set standards in 1992. The methods used student work, teacher evaluations of classroom performance, and difficulty rankings of actual test items to set the standards. A final step synthesized the varying results from the three methods into Kentucky's standards that are hoped to be stable for many years. Contrary to some critic's claim that the new standards turned CATS into a norm-referenced test, this is not the case. The new cut-points or standards are clearly tied to the Core Content and to specific points that students must achieve, regardless of the percentage of students that achieve that category. An additional result of the new standards was the creation of a set of clear definitions of what each performance level represents, definitions useful to both teachers and parents.

WHAT IS THE DIFFERENCE BETWEEN A STANDARDS-BASED TEST AND A NORM-REFERENCED TEST?

As indicated above a standards-based test expects students to reach a certain level on the test to reach a category. It does not matter how many students achieve the standard. The mark remains the same. In Kentucky, at present, more students are in the bottom two categories than are in the top two. The goal is to reverse that situation by 2014. For a norm-referenced test, students are assumed to follow a certain curve with 68% of the students within one standard deviation on either side of the mean, and approximately 95% within two standard deviations on either side of the mean. If students begin to increase their scores the test has to be re-normed to remain useful. That means the target for the student moves as scores improve, whereas the target for the student remains stable, and therefore known to all, in a standards-based test.

WHO GIVES THE TEST?

Once the forms have been constructed, they are shipped to a second contractor, currently Data Recognition Corporation of Maple Grove (Minneapolis), Minnesota. There the test booklets, and answer booklets are printed, quality checked, boxed by school and shipped to the 176 school District Assessment Coordinators. These administrators at the local level check the boxes to make sure each school has adequate materials, and distribute the tests to the schools a few days before the testing window (around late April and early May). The school has a Building Assessment Coordinator who is responsible for making sure that teachers who administer the test follow instructions. Some students may take the entire test over several subjects in one location. Others may be in a different room each day. Schools have several different patterns they may follow regarding how much testing is done each day. The crucial issue is that all students at a grade level must do the same sections of the test on the same day. The Kentucky Department of Education provides Administration Manuals for teachers to use that tell them exactly how to give the test and exactly what to say, so that all students have an equal chance to do well.

HOW DO WE KNOW THE TEST WAS GIVEN FAIRLY?

The main safeguard of fairness is the integrity of Kentucky teachers. While we sometimes read in the papers about a teacher doing something illegal, the fact is that teachers are among the most honest and truthful groups of people in the state. Even if you have had a bad experience with a teacher, that does not necessarily mean they are dishonest or untruthful. In addition, there is an "allegation" process where parents, teachers, or administrators can file a complaint or an admission if something was done incorrectly. A division of KDE that is completely separate from the Office of Assessment and Accountability investigates the allegations. If the allegation proves true, it may fall into one of two categories. One includes those incidents that do not affect student scores. The other category includes those allegations that do affect student scores. Student scores may be changed to zero, which punishes the school for not administering the test appropriately. It should be noted that in these cases, parents still receive a score report for their children that has original scores, but a zero score is used for purposes of school accountability at the school level.

The number of allegations per year ranges from 100 to 200. In light of the more than 30,000 teachers who administer tests each year, this is a very small amount. Test scores change for a few hundred students each year of about 400,000 tested each year.

WHAT ABOUT PORTFOLIOS?

Kentucky is one of the few states that have a statewide portfolio requirement that is used to aid in evaluating schools. The submission of writing portfolios occurs in grades four, seven and twelve. Work on the pieces submitted may take place at any grade level. Students submit a specified number of pieces that exhibit ability to complete different kinds of writing like personal narratives, persuasive, or practical workplace writing. One piece must be from a subject other than language arts. The portfolios are scored at the school according to specific requirements (called rubrics) by groups of teachers, language arts teachers at some schools, and all teachers at others. Each year KDE in cooperation with a contractor conducts an audit of 100 schools: fifty selected randomly, and 50 selected because they exhibited a large change in scores. The accuracy of scoring is verified for these schools.

WHAT IS AN ALTERNATE PORTFOLIO?

In a prior question we mentioned steps taken to make it possible for students that are impaired to have an equal chance to perform well. There are, however, some students so severely impaired intellectually or physically that they are unable to perform with a paper or pencil test. In Kentucky, somewhat less than one percent of the students fit in this category. A special means has been developed to measure the progress of these students, called the alternate portfolio.

Each student with a severe impairment is in a classroom with fewer students, although they may spend some of their day in a regular classroom, with modified assignments, and with the help of a supporting person. They have an individualized plan of educational goals, which are selected from Kentucky's Academic Expectations and the *Core Content for Assessment*.

The Alternate Portfolio is the tool for assessing progress toward the goals selected for the student. The required contents of the portfolio include a table of contents, a student letter to the reviewer, a parent letter validating the portfolio, the student's schedule, a summary of job exploration at grade 8 or a resume at grade12, and five entries which represent the required subject areas at the student's grade level.

Alternate portfolios are scored by two teams of two teachers who are familiar with the construction of alternate portfolios. Scoring takes place at the regional level. Agreement between the teams makes the score final. Disagreement leads to scoring by a state expert whose decision is final. A single category (novice, apprentice, proficient or distinguished) is given to the portfolio. In the accountability index for the school, the student score counts in each subject area required at that grade level.

VALIDITY AND RELIABILITY

DOES THE TEST MEASURE WHAT IT IS SUPPOSED TO?

Validity is the appropriateness, meaningfulness and usefulness of the conclusions drawn from test scores. KDE takes very seriously the standards of national professional organizations relating to validity. Careful data is maintained about both the teachers who write the questions and about the match between the Core Content and the KCCT. When the teachers write the questions, they assign a primary and possibly a secondary Core Content "bullet" that the question is intended to measure. The contractor's experts evaluate these assignments and give feedback to the Content Advisory Committee if they disagree. The question is then reconsidered by the CAC. Problematic questions may never make it to the test, but if the question is regarded as exceptional, it may be tested. Research concerning how students answered each experimental question (i.e., each pre-test question) may well enlighten the CAC regarding whether these questions allowed the desired response from students.

A second means of checking whether the test measures what it is supposed to is an annual report of all the assigned Core Content codes on the test. This report is used to see if the test is properly balanced and covers all the content bullets in the six forms in a subject area. This report is compared to a document called the *KCCT Blueprint* that specifies the percentage of questions on the test for each subdomain. Kentucky teachers created the Blueprint, with the help of KDE staff. The annual report indicates whether the percentages specified in the Blueprint are being met. The report guides the writing of questions to specific topics where there may be a gap, and also guides the form building process.

One of the most important issues with regard to any test is whether the test considers appropriate criteria such as cognitive complexity (how hard the questions causes a person to think) or content quality (how well the question measures the content). One way of looking at this question is whether the student who answers the questions can show proficient and distinguished performance. This issue has been carefully addressed by the CACs. The setting of new standards also has an impact upon this question (See below). Another way of answering this question is whether high scoring schools do things differently than low scoring schools. Several studies have accumulated, as well as results from school audits that indicate that high scoring schools are very intentional in aligning their curriculum to the Core Content, have rich and rigorous curricula, and have aligned classroom assessment with the types of assessment that appear on the KCCT.

DOES THE TEST MEASURE RELIABLY?

Since many who ask this question are referring more to whether the test is accurately scored, than to formal reliability, we will consider that separately below. With regard to the reliability of the KCCT, in reading, mathematics, science and social studies at most levels, the reliabilities are between .80 and .89 which is excellent. For the shorter tests in arts & humanities and practical living/vocational studies, as expected the reliabilities are lower. They were .60 to .69, which is acceptable. For more information concerning reliability see the CATS 2000 Technical Report.

IS KENTUCKY'S TEST FAIR TO ALL STUDENTS?

We have mentioned some fairness issues in earlier questions. We briefly discussed some of the means of allowing impaired students to have an equal chance to succeed (these are called accommodations). We also considered fairness in administering the test. The most frequent fairness concerns involve gender and race. There is a consistent pattern over the years of girls outperforming boys in language arts and social studies at the middle and high school levels. There is a second pattern of boys performing better than girls in mathematics at the high school level. There is a consistent pattern in the test results of those with an Asian heritage outperforming all students, and of Caucasians outperforming African-Americans and Hispanics. Do these results reflect bias in the test or are they an accurate reflection of the results of the educational process?

Kentucky uses two methods to make sure that such performance differences are not due to bias in the test. The first means is the Bias Review Committee (BRC). This group, which represents a broad cross-section of educators, business people, and special concern groups, meets twice annually. In the spring this group reviews reading passages that will be used by the CACs to write questions. The BRC looks for concepts that are only known to a few at the grade level, things that might offend or distract students from a racial, religious or social group, things that are outside the experience of a social grouping, or passages that do not lend themselves to use by the blind or hearing impaired. The fall meeting of the BRC is spent reading the actual questions that will be considered for experimentation for the same kinds of bias mentioned above.

The second method of finding bias is quantitative, that is, it is based on mathematical analysis. The method is called Differential Item Functioning (DIF). This method compares how the item worked in comparison to all the other items on the test of like kind. If an unusual pattern for an item is discovered between groups of students, it may be sent back to the BRC to be rechecked for bias, or the question may be removed from the test. Kentucky uses the most elaborate and complex method available for checking DIF, and possible bias.

If the test is not biased, then what explains the performance differences between groups? This becomes an instructional issue. Are girls **expected** to do as well as boys in mathematics? Are African-American students **expected** to do as well as Caucasian students. The Instructional Equity team of KDE, as well as the Division of Equity address equality of opportunity and of expectations. The issue of equal opportunity and equal expectations is also a component of the audit process for low performing schools.

WHO CHECKS TO MAKE SURE THINGS ARE DONE RIGHT?

We have already answered this question in part. The contractors and the Bias Review Committee check the work of the Content Advisory Committees. In turn, KDE checks the work of the contractors. In past years, teams from the Division of Assessment Implementation and the Division of Validation and Research visited each site where Kentucky tests are built, printed, scored and reported. Specific points of concern were identified in advance and carefully reviewed by KDE staff on these visits. In addition, there are many advisory groups that assist in making Kentucky's test one of the best in the land. A group of nationally recognized testing experts advise KDE on technical issues. This group, the National Technical Advisory Panel for Assessment and Accountability, or NTAPAA, meets quarterly. There are other in-state groups of advisors representing teachers, principals, superintendents, school boards, parents, professional groups, business people, chambers of commerce, the legislature, and others. New plans are passed before these groups before the Kentucky Board of Education (KBE) takes action. The KBE has the ultimate responsibility for making sure things are done right. Despite what a few vocal critics might say, attendance at these meetings soon demonstrates that Kentuckians are dedicated to building both the best test possible, and an educational system that is successful.

SCORING

HOW DO WE KNOW THE TEST IS GRADED FAIRLY?

Kentucky assures fairness in scoring the Kentucky tests by contracting with independent contractors who have no vested interest in the outcome. The contractor is experienced in scoring and has many checks and rechecks built into the scoring system. As an example, the most experienced scorers reread 2% of all the questions to make sure that the original scorer is on track. The papers are randomly selected and the original scorer never knows which ones will be read. This is called a double read process. A second method is that scorers are organized into teams of ten with a leader. Once a day the leader reads approximately ten papers from each of his/her ten scorers. This represents 7 to 10% of each scorer's daily production. If a scorer has strayed they are immediately put back on track, and all the papers they scored that day may be re-scored. Kentucky requires an 80% perfect agreement rate for scorers to qualify to score Kentucky papers. This is the highest requirement of any of the states served by the current contractor.

CAN OPEN-RESPONSE REALLY BE SCORED CONSISTENTLY?

While 80% perfect agreement between scorers doesn't sound very good initially, it becomes more impressive when we realize that it is higher agreement than is common for classroom essays. It is also easier to accept when we realize that the student does not suffer any consequences if his answer is scored incorrectly, if only one of several questions is off by one point. At the school level the questions that were incorrectly scored down are somewhat compensated by those incorrectly scored too high.

WHO SCORES KENTUCKY'S TEST?

The contractor hires those who score the tests. Kentucky requires that all scorers have at least two years of college, however, over 90% of the scorers have a college degree and many have advanced degrees, especially teachers and retirees. A sizable number of scorers at the six or seven sites that score Kentucky papers are teachers, but many other professions are represented as well. Teachers do not always make the best scorers, because some cannot accept the Kentucky rubric (scoring guide) without challenging it. This is important because KDE obviously wants the scoring to go according to the Kentucky designed and built rubric. The current contractor has scoring sites in Minnesota (at least five locations), Chicago, Cincinnati and Wilmington, North Carolina. The ethnic composition of scorers is approximately 13% minorities, which closely matches Kentucky's 15% minority student population. More females score than males, but then there are more female teachers in Kentucky schools.

WHAT GUIDES SCORING?

The most significant piece of the process of accurate scoring, however, is the care with which the scoring guide (rubric) for a given question is written. The CAC member that writes the question also writes the scoring rubric. Using the rubric, they describe what student work will look like for each of the score points assigned. Most questions have either four or five total points possible, and the rubrics often specify how half points can be achieved. At the end of the rubric all scores are converted to a standard four-point scale with no half points. Kentucky rubrics have drawn praise from the contractor's readers for their completeness and ease of use. The experienced scorers used by the contractor become very capable of making consistent decisions about student papers hour after hour and day after day.

REPORTING

WHY DOES IT TAKE SO LONG TO GET THE RESULTS?

Scoring essay type questions for a whole state (well over 400,000 students) takes time. Just the packing at schools and unpacking at the contractors with the checking in of every paper takes several weeks. Two or three days are necessary to score the six questions on one form. There are six forms, and six subjects with the multiple forms, and the writing test, all of this at three different levels. So, the result is two months or more just to score. Then all the statistical work must be done to produce the information for individual students, the schools, the districts and the state. Even the simple printing and shipping of reports takes a great deal of time. So the time from the end of May to mid September turns out to be short in terms of producing what Kentucky needs. A simple multiple-choice test could give us quicker data, but a lot less information about how Kentucky schools are achieving.

WHAT IS AN ACCOUNTABILITY SYSTEM?

One of the most confusing aspects of Kentucky's testing system is the difference between the KCCT and CATS. The first is the actual test. The second is the accountability system that in fact includes the CTBS/5 and KCCT plus other indicators of school performance. The objective of CATS is to have the same goal for all schools, proficiency by 2014. But schools are starting at different points. Some schools are already excellent, but some are not.

CATS is designed to measure progress toward the goal. Simply put, a starting point was established for every school during the 1999 and 2000 biennium. The new standards were applied to the scores for those years to establish the starting point. Proficiency is defined as a score of 100 on a 140-point scale. A line is drawn from where the school was in 1999-2000 to a score of 100 in 2014. This creates a chart with a line connecting two points, which is called the goal line. Schools whose score (or accountability index) is at or above the line are meeting the goal and are eligible for financial rewards.

In addition to the goal line, a second line is drawn from the 1999 and 2000 biennium point to an index or score of 80. This line is called the assistance line. Schools in between the two lines are "progressing" if their scores are increasing. These schools are eligible for smaller rewards. If the school scores remain the same or declines the school receives no rewards. Schools below the assistance line undergo a state review, and those in the bottom third of schools below the assistance line are audited to determine what financial and professional help they need to improve.

Because there is always the possibility of measurement error in any type of scoring (all tests have at least some measurement error), the goal line is actually drawn to a point slightly below 100 to take this possible error into account. Similarly, the assistance line is drawn to a point slightly below 80 to take possible error into account.

In addition to the above rewards system, for schools that are improving there are five recognition points where additional rewards may be earned. Also, the top five percent of schools, if they are above the fourth recognition point, may be designated Pacesetter schools.

HOW DO REWARDS AND ASSISTANCE WORK?

Each year the Kentucky Board of Education determines the amount of money available for rewards. Information is gathered on the number of schools and the number of teachers in those schools in order to calculate the value of a share of rewards. Schools that are above their goal line, and have met their novice reduction and/or dropout goals, are eligible for three shares of rewards. Improving schools between the goal line and the assistance line receive one-half share of rewards. Schools that exceed for the first time one of five recognition points receive one share of rewards. "Pacesetter" schools that are past the fourth recognition point and that have not declined in the previous two biennia, and that are in the top 5% of schools and have met their novice reduction target, are entitled to a share of rewards. The dollar amount the school receives is the number of shares it is entitled to times the number of full- time teachers. School councils decide how to spend reward money and may choose from several options including materials, supplies or bonuses for teachers and other staff.

For schools below their assistance line, some or all of the following forms of assistance may be received: an invitation to draft a school improvement plan, a scholastic audit to recommend specific assistance needed, Commonwealth School Improvement Funds, a highly skilled educator to provide advice, and an evaluation of school personnel. The goal of assistance is to aid the school in beginning the process of making continuous improvement toward proficiency by 2014.

WHAT IS NAEP AND WHY SHOULD I CARE?

NAEP means the National Assessment of Educational Progress, and is frequently called "the nation's report card." Tests are given to a sample of students every four years in a subject. Currently reading, mathematics, writing and science are tested. Kentucky has participated in this testing program since the beginning of state level testing in 1990. Results over the past decade show that Kentucky has been improving in the tested subjects in reading and math at the 8th grade level. Kentucky students are drawing near to the national average in both 4th and 8th grades. NAEP serves as a partial check that the progress made on the CATS is real and genuine. Other national programs like the ACT also provide some evidence. A growing number of high school students are taking the ACT. Normally when a larger number take the test, the assumption is made that scores will go down because a larger number of less able students are taking the test. The good news for Kentucky is that as the number of students taking the test has grown, scores have remained relatively the same.

HOW DO I GET THE KCCT RESULTS FOR MY CHILD?

In the fall of each school year in mid-September each school receives the scores of students who participated in testing the previous spring. These scores for individual students are sent to parents or guardians a short time after the school receives them. The scores for CTBS/5 are received in mid-August, but are sent to parents after the beginning of the school year in most cases. If you do not receive scores for your child who is in the 4th through 12th grade call your school.

HOW DO I LEARN ABOUT MY CHILD'S SCHOOL?

In addition to a report for the individual student, the school is also required to produce a document called the School Report Card. This document tells important information about the school, and the school's performance on the CATS, attendance rates, how many children were held back for a second year in a grade, how much money is spent per student, parent participation, the percentage of teachers with degrees in what they teach, the percentage who have a master's degree, and many other topics. A printed version of the School Report Card is sent home to parents by mid-January. The School Report Card for each school can also be viewed on the Kentucky Department of Education website.

OTHER MATTERS

WHO DOES NOT HAVE TO TAKE THE TEST?

Less than one percent of Kentucky students are excused from the test. These students include:

- Students who move out-of-state before the testing window.
- Students with a medical condition that prevents them from taking the test may be exempted on the basis of a doctor's recommendation and concurrence by KDE. It should be noted that many medical disabilities are accommodated by means of a scribe or a computer. Students with Individual Educational Programs may also receive accommodations and be able to complete the test. Some students do the alternate portfolio.
- Another group that does not take the test are those who have dropped out or graduated before the test date.
- Students who have not been enrolled in Kentucky schools enough days may be exempted from completing the Writing Portfolio.

- Students who are English Language Learners may also be exempted from the test.

As can be seen, Kentucky makes every effort to test every student who possibly may be fairly tested.

WHAT ARE NONACADEMIC INDICATORS?

Ten percent of a school's accountability score is based on "nonacademic indicators." These are items that are not subject matter oriented but are very important to success. Included in these are the school's percentage of attendance, the percentage who are required to repeat a grade, the percentage who drop out of school during grades 7 through 12, and at the high school the percentage of students who make a successful transition to adult life. This successful transition is demonstrated by such things as becoming employed, joining the armed forces, entering college or a vocational school, and others. Some of these items must be based on data gathered a year earlier than the testing year in order to be complete.

APPENDIX

(THINGS YOU MAY OR MAY NOT WANT TO KNOW)

TESTING THE LEARNER GOALS

Kentucky has six goals for learners, established by law.

KENTUCKY'S SIX LEARNER GOALS

1. Students shall use basic communication and mathematics skills for purposes and situations they will encounter throughout their lives.
2. Students shall develop their abilities to apply core concepts and principles from mathematics, the sciences, the arts, the humanities, social studies, practical living studies, and vocational studies to what they will encounter throughout their lives.
3. Students shall develop their abilities to become self-sufficient individuals.
4. Students shall develop their abilities to become responsible members of a family, work group, or community, including demonstrating effectiveness in community service.
5. Students shall develop their abilities to think and solve problems in a variety of situations they will encounter in life.
6. Students shall develop their abilities to connect and integrate experiences and new knowledge from all subject matter fields with what they have previously learned and build on past learning experiences to acquire new information through media sources.

Goals three and four are not tested by the KCCT because it is difficult to devise meaningful ways of evaluating these and the evaluation could raise issues of personal privacy.

DISTRIBUTION ACROSS ACADEMIC EXPECTATIONS

Assessing the quality of the KCCT includes making sure that the test is properly and comprehensively related to the 57 Academic Expectations. At least once per biennium tables are produced that demonstrate the distribution of items across the Academic Expectations. These tables are not included here, but are available in the Technical Reports produced by the Office of Assessment and Accountability and the contractors. These are available upon request.

DISTRIBUTION ACROSS CORE CONTENT

In a fashion similar to distributions across the Academic Expectations, tables of distribution of items with regard to the Core Content are produced annually. These are carefully checked to make sure the test is matching the Blueprint and to provide guidance to the contractor during the building of the six forms in each of the subject areas tested.

ITEM ANALYSIS

To provide evidence of the technical quality of the KCCT a series of item level analyses are performed for each grade and subject area. The following list summarizes some of the analyses conducted.

- Distribution of item scores for open-response items,
- Distribution of corrected item-total correlations for open-response items,
- Distribution of item-theta correlations for open-response items,
- Distribution of N vs. A, P, D biserial correlations for open-response items,
- Distribution of N, A vs. P, D biserial correlations for open-response items,
- Distribution of N, A, P vs. D biserial correlations for open-response items.

A comprehensive overview of the above analyses, which is an enormous amount of statistical data, is available in the various Technical Reports about the KIRIS and CATS systems. These reports are available from the Office of Assessment and Accountability, Kentucky Department of Education.

In 1998 some initial work on differential item functioning (DIF) was begun. The purpose of these studies was to determine if items function differently for subgroups of students, such as males versus females, or African Americans versus Caucasians. While DIF is a requirement for bias to be present, it is not sufficient to indicate bias, which has to be addressed by the Bias Review Committee. A much larger project across all grade levels and in all subject areas was completed in 2001 and a summary of the results is available from the Office of Assessment and Accountability, Kentucky Department of Education.

SCALING

Scaling is the process of making sure that a score on one form of the test means the same thing as a score on a different form. Scaling is also necessary from year to year to make sure that a given score means the same thing year after year. Scaling involves converting raw scores into scale scores (Kenctucky uses a scale from 325 to 800) and doing statistical processes that establish the desired comparability. For more information see the Technical Reports to which we previously referred.

PORTFOLIO AUDITS

Each year 100 schools are selected to participate in a writing portfolio audit. The purpose is to verify that scoring is being done accurately at the school site since the contractors do not score portfolios. Fifty of the schools are selected randomly, although this sample is divided into three groups: elementary, middle and high school. The other fifty schools are selected on the basis of having the greatest amount of change in their portfolios scores, either up or down. The purpose here is to make sure the changes are real. Once a school is selected it sends all its portfolios to the contractor, where professional scorers evaluate the portfolios using the same rubric (scoring guide) that the teachers used. The scores given by the contractor are the final scores. Many schools are right on target. Some schools grade too easily and some schools grade too hard. In the past, each fall the audited schools would meet with KDE staff during which means of improving the portfolios and means of improving the accuracy of the scoring were discussed.

SCHOOL AUDITS

A school audit, or scholastic audit, normally happens when a school does not meet its goal, and scores among the bottom third of the schools that fell below the assistance line. A diverse team of five educators, the make-up of which is established in regulation, visits the school for a week. Each member of the committee has particular responsibilities: data review, budget analysis, classroom visitation, administrative evaluation and other issues. The analysis takes place during four or five days and a lengthy report is written. The guide for the visiting team is called the Standards and Indicators for School Improvement or SISI. This document organizes a school in terms of its success in meeting ten specific standards for excellence. The indicators are specific evidence that the team looks for that indicate the level of functioning of the school. The evaluation for each indicator is on a scale of one to four and provides discussion of specific reasons for the category into which the school is placed. The intention of the audit is to give the school specific guidance concerning its weaknesses that are causing the students to fail to perform better.

HISTORICAL TIMELINE

The following is a brief summary of actions related to Kentucky's system of assessment and accountability. These are actions taken by the Office of Assessment and Accountability and its predecessor, Office of Curriculum, Assessment, and Accountability.

1990 The OAA assisted NAEP in the 1990 8th grade reading assessment.

Technical assistance was elicited for psychometric advice from experts in the field. The National Technical Working Group (later the National Technical Advisory Panel for Assessment and Accountability) was formally established in 1995.

1991 The OAA assisted in the gathering of information for drafting the 75 Academic Expectations (originally referred to as Valued Outcomes).

1992 The OAA assisted with the drafting of the first set of performance standards.

The OAA, in conjunction with contractors, constructed, administered, scored and reported the first KIRIS assessment for the purpose of establishing baselines for the accountability system for schools.

The first teacher groups (later Content Advisory Committees) were formed to participate in writing and selecting the questions for the KIRIS assessment.

In the following years the KIRIS and its successor CATS used a wide variety of assessment types for the purpose of validity, accuracy of assessment and assisting in modifying instruction including multiple-choice (Pre-tested in the spring of 1997 and 1998, and entered "Long-Term" accountability in 1999), open-response, performance events (1993 to 1996), portfolios (Writing all years, mathematics, 1993 to 1996), and on-demand writing.

The OAA supervised through a contractor the administration and scoring of the alternate portfolio, which was included in the accountability system beginning with the 1992-1993 school year.

The OAA assisted NAEP in the 1992 assessment of 4th grade reading and mathematics, and 8th grade mathematics.

Beginning in 1992, item level reporting was begun to improve student motivation. Changes have been made incrementally from 1992 to 2002 to improve the process.

1993 The OAA provided through a contractor the first technical manual with detailed information concerning the assessment.

The OAA provided the first professional development for the District Assessment Coordinators, and provided the first Implementation Guidebook.

The OAA with assistance from the contractors conducted the first audit of Portfolio scores. After scoring accuracy analyses conducted in 1994 and 1995, the audits became a regular feature.

KIRIS Curriculum and Assessment Reports were initiated for purposes of accountability. These later became the KIRIS Performance Reports (1997) and the Kentucky Performance Reports (1999).

1994 The OAA adjusted the assessment process based on the legislative withdrawal of Learner Goals 3 and 4 from assessment, and aided the reformulation of the 75 Valued Outcomes into the 57 Academic Expectations.

The OAA again assisted NAEP in the assessment of 4th grade reading.

The first KIRIS cycle ended with the assignment of rewards and sanctions.

The OAA assisted in the establishing of the first *Content Guidelines*.

The OAA assisted with the production of the portfolio implementation manuals.

1995 The OAA assisted in the study/validation of the 1992 performance standards.

1996 First *Core Content for Assessment* document produced (Revised by Curriculum Division in 1999).

The OAA assisted NAEP in the administration of assessments in 4th grade mathematics and 8th grade mathematics and science.

Assessment Cycle 2 ends with appropriate rewards and sanctions.

1997 The administration of the CTBS/5 Survey Edition began.

1998 The third KIRIS cycle ended with the assignment of rewards and assistance.

The OAA assisted with the NAEP assessments in 4th grade reading, and 8th grade reading and writing.

1999 The CATS Interim Cycle begins.

A two-year multi-step standard setting project was initiated.

The CTBS/5 Survey Edition is included in the Long-Term Accountability index.

The Validation and Research Division was initiated. The OAA has engaged in a continuously expanding program of validation over the decade, assisted by the contractors and focusing primarily upon construct and consequential validity, which led to the creation of the division.

2001 The standards setting process was completed and the new standards for the KCCT were adopted by the Kentucky Board of Education.

Goal lines and assistance lines were recalculated for all schools and growth charts produced based upon the new standards.

The Longitudinal Accountability Pilot Project continued.

A major project to assess Differential Item Functioning (DIF) was initiated to determine if any items were potentially discriminating against a subgroup.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

NOTICE

Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (1/2003)